# NAVAL POSTGRADUATE SCHOOL
## Monterey, California

# THESIS

US ARMY'S DELAYED ENTRY PROGRAM:
ATTRITION MODELING

by

Daniel C. Buning

September, 1991

Thesis Advisor: Donald R. Barr

Approved for public release; distribution is unlimited

T257740

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

# REPORT DOCUMENTATION PAGE

| | |
|---|---|
| 1a REPORT SECURITY CLASSIFICATION<br>Unclassified | 1b RESTRICTIVE MARKINGS |
| 2a SECURITY CLASSIFICATION AUTHORITY | 3 DISTRIBUTION/AVAILABILITY OF REPORT<br>Approved for public release; distribution is unlimited. |
| 2b DECLASSIFICATION/DOWNGRADING SCHEDULE | |
| 4 PERFORMING ORGANIZATION REPORT NUMBER(S) | 5 MONITORING ORGANIZATION REPORT NUMBER(S) |
| 6a NAME OF PERFORMING ORGANIZATION<br>Naval Postgraduate School<br>6b OFFICE SYMBOL (If applicable)<br>OR | 7a NAME OF MONITORING ORGANIZATION<br>Naval Postgraduate School |
| 6c ADDRESS (City, State, and ZIP Code)<br>Monterey, CA 93943-5000 | 7b ADDRESS (City, State, and ZIP Code)<br>Monterey, CA 93943-5000 |
| 8a NAME OF FUNDING/SPONSORING ORGANIZATION<br>8b OFFICE SYMBOL (If applicable) | 9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER |
| 8c ADDRESS (City, State, and ZIP Code) | 10 SOURCE OF FUNDING NUMBERS<br>Program Element No / Project No / Task No / Work Unit Accession Number |

11 TITLE (Include Security Classification)
US Army's Delayed Entry Program: Attrition Modeling

12 PERSONAL AUTHOR(S) Buning, Daniel Christopher

| 13a TYPE OF REPORT | 13b TIME COVERED | 14 DATE OF REPORT (year, month, day) | 15 PAGE COUNT |
|---|---|---|---|
| Master's Thesis | From To | 1991, September, 26 | 78 |

16 SUPPLEMENTARY NOTATION
The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

17 COSATI CODES

| FIELD | GROUP | SUBGROUP |
|---|---|---|
| | | |
| | | |

18 SUBJECT TERMS (continue on reverse if necessary and identify by block number)
Delayed Entry Program, United States Army Recruiting Command, attrition, DEP Loss

19 ABSTRACT (continue on reverse if necessary and identify by block number)

The United States Recruiting Command (USAREC) utilizes the Delayed Entry Program (DEP) as the foundation for their management of the continous flow of recruits into the training base. Though there are many benefits of the DEP, a major shortcoming is that some DEP members do not enlist, becoming DEP losses. This is costly in terms of valuable resources such as lost recruiter time, and the potential for training seats being unfilled. Any effort which assists in reducing DEP loss would be a valuable contribution. This research models individual level DEP loss using multivariate dichotomous logistic regression. Explanatory variables used were individual, demographic, and USAREC policy in nature. Modeling efforts used data that were easily accessible to USAREC to ensure ease of potential future use. Univariate analysis was conducted on canidate explanatory variables prior to model building. The model was built using forward and backward stepwise logistic regression. Final model refinement included scaling of interval variables and the addition of one interaction term. Using statistical tests, the model as a whole was determined to exhibit some lack of fit. Closer analysis indicated that the model does perform well accross many levels of estimated probability of DEP loss. Using USAREC's red, amber, green DEP loss risk classification system, the model appears to have significant predictive powers. The model also performed well using this classification system for a validation data set. It is concluded that this fitted model could prove useful in supplementing the field experience of the recruiter in predicting DEP loss risk of individual recruits.

| | | |
|---|---|---|
| 20 DISTRIBUTION/AVAILABILITY OF ABSTRACT<br>☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS REPORT ☐ DTIC USERS | 21 ABSTRACT SECURITY CLASSIFICATION<br>Unclassified | |
| 22a NAME OF RESPONSIBLE INDIVIDUAL<br>Donald R. Barr | 22b TELEPHONE (Include Area code)<br>(408) 646-2763 | 22c OFFICE SYMBOL<br>MA/Ba |

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

Approved for public release; distribution is unlimited.

US Army's Delayed Entry Program:
Attrition Modeling

by

Daniel C. Buning
Captain, United States Army
B.S., United States Military Academy, 1982

Submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1991

Peter Purdue, Chairman
Department of Operations Research

## ABSTRACT

The United States Recruiting Command (USAREC) utilizes the Delayed Entry Program (DEP) as the foundation for their management of the continuous flow of recruits into the training base. Though there are many benefits of the DEP, a major shortcoming is that some DEP members do not enlist, becoming DEP losses. This is costly in terms of valuable resources such as lost recruiter time, and the potential for training seats being unfilled. Any effort which assists in reducing DEP loss would be a valuable contribution.

This research models individual level DEP loss using multivariate dichotomous logistic regression. Explanatory variables used were individual, demographic, and USAREC policy in nature. Modeling efforts used data that were easily accessible to USAREC to ensure ease of potential future use. Univariate analysis was conducted on candidate explanatory variables prior to model building. The model was built using forward and backward stepwise logistic regression. Final model refinement included scaling of interval variables and the addition of one interaction term.

Using statistical tests, the model as a whole was determined to exhibit some lack of fit. Closer analysis indicated that the model does perform well across many levels of estimated probability of DEP loss. Using USAREC's red, amber, green DEP loss risk classification system, the model appears to have significant predictive powers. The model also performed well using this classification system for a validation data set. It is concluded that this fitted model could prove useful in supplementing the field experience of the recruiter in predicting DEP loss risk of individual recruits.

Thesis
B852
c.1

TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

# I. INTRODUCTION

The United States Army Recruiting Command (USAREC) utilizes the Delayed Entry Program (DEP) as an important management tool in ensuring the US Army receives a continuous flow of recruits. The Delayed Entry Program provides benefits to the recruit and the Recruiting Command alike. A major shortcoming of this program is that some newly contracted recruits in the DEP pool do not enlist. This attrition process is costly in recruiting resources and potentially results in training seats being unfilled. This research models the DEP loss process in an attempt to identify contracts with relatively high risks of DEP loss.

## A. DELAYED ENTRY PROGRAM DESCRIPTION

The DEP is an enlistment program which allows an individual to delay entry onto active duty for a period of up to 365 days. It is best thought of as a reservation system. Qualified applicants are allowed to contract for enlistment at a specified time, for particular training, and a guaranteed job, for an agreed upon time of service [Ref. 1]. The recruiter keeps in close contact with the DEP member to help ensure that he remains mentally and physically qualified for enlistment, and that he maintains his desire to enlist. DEP management is any activity that promotes this accession goal

and includes funded and unfunded DEP functions, optional military training or instruction, and other activities. DEP management is quite similar to the initial recruiting process in that the initial contract is continuously resold while the recruit is in the DEP. [Ref. 2]

The day in which a young person could walk into the Recruiting Office, sign up, and ship out is gone. With the arrival of the Drug and Alcohol Testing (DAT) in June 1988, DEP is the vehicle in which all recruits enter active duty.

## B. DEP BENEFITS

The DEP provides benefits for both the recruit and USAREC. The DEP allows the recruit to lock in training, schooling and an assignment, many months in advance. A recruit in high school can make definitive plans for the future early in his senior year. The DEP also allows the recruit a wider range of available assignments. The recruiter is able to project out one year for available assignments. This is especially valuable for the top quality recruit who qualifies for all assignments.

The DEP provides benefits to the US Army because it allows for efficient resource management in a business that tends to be extremely seasonal. The DEP aids in future planning of training availability and personnel requirements. Recruiters are able to focus on high quality recruits rather than meeting short term accession goals. US Navy research efforts indicate

that a large DEP pool may actually assist recruiting [Ref. 1]. This may be due to the promotion incentives offered to DEP members who refer candidates who then enlist. In effect, every DEP member becomes a recruiter, representing the US Army in the high schools and work places, creating a type of recruit network.

Another byproduct of the DEP is that it may result in lower first term attrition. One study conducted for the US Army in 1985 concluded that the longer the recruit was in the DEP the more likely he was to successfully complete his term of service. The theory of this study is that a recruit who has more time to evaluate his contract decision, and then accesses onto active duty, will be more inclined to fulfill his contractual obligation [Ref. 3]. A related theory is that someone who survives a longer period in the DEP may be more committed to begin with, so that a portion of the total attrition occurs in the DEP rather than after enlistment.

## C. DEP SHORTCOMINGS

The DEP is not without its costs to USAREC. During the period a recruit is in the DEP, he may attrite or become a DEP loss. A DEP loss may be the result of a myriad of reasons ranging from death or serious injury, to apathy, to joining another service or National Guard. During the last ten years, DEP loss has grown from 7% upwards to 13% in FY 89. As of 1 December 1990, approximately 15% of all contracts signed in FY

90 resulted in DEP losses.[1] Figure I depicts the trend over the last 20 quarters. Large DEP losses significantly contributed to USAREC not meeting its accession goals in October and November 1990, the first time in over seven years.

USAREC Regulation 601-95 states, "DEP loss has a major impact on mission accomplishment." A DEP loss must be replaced by a new recruit, demanding valuable recruiter resources and time. If a DEP loss occurs shortly before the accession date, a training seat could remain unfilled. With smaller defense budgets, the US Army cannot afford to under utilize its training resources. In the last year, USAREC reports that recruiters are finding they must make on the average 12 contacts with potential recruits, versus an average of 8 in previous years, to secure one enlistment [Ref. 4]. This indicates that it may become even more difficult to recruit replacements for DEP losses in the future.

## D. CURRENT USAREC DEP SYSTEM

USAREC's command goal is to reduce DEP loss to six percent or less of all signed contracts [Ref. 2]. As Figure 1 indicates, this goal has not been reached in any of the last 20 quarters and only during two, one month periods in FY 90. USAREC Regulation 601-95 outlines many approved techniques to

---

[1] As of 1 December 1990, approximately 80% of all contracts signed in FY 90 had resulted in accessions or DEP losses. The remaining recruits were still awaiting accession onto active duty or DEP loss.

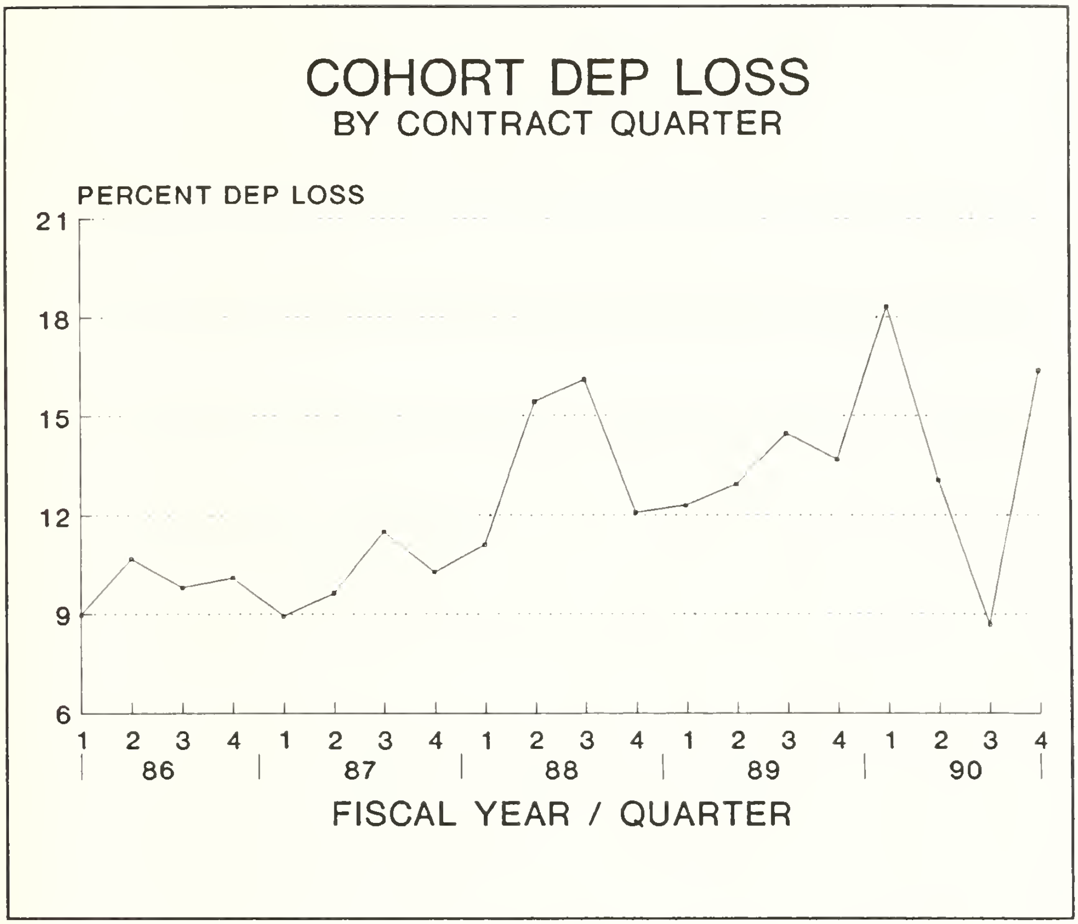


Figure 1 Cohort DEP Loss by Quarter FY86 - FY90

help avoid DEP losses. These include: minimum standards for number of times a recruiter contacts a DEP member, DEP incentive programs, and funded DEP events. Currently, recruiters rely only on their experience in the field to categorize their recruits in the DEP as being high, medium, or low DEP loss risks. Recruiters are required to report to their chain of command monthly their subjective opinion as to the risk status of their DEP members using the following coding scheme:

- Green: Indicates the DEP member remains motivated to access onto active duty and there are no foreseeable problems.
- Amber: Indicates there may be potential problems with either motivation or qualification to access onto active duty.
- Red: Indicates a problem. This DEP member for whatever reason is a probable or certain DEP loss.

This system of using the field expertise of the recruiter and his personal knowledge of each DEP member appears to be valuable. USAREC could potentially augment this system with quantitative techniques or models to better assist in predicting DEP losses.

Chapter II summarizes the goals of this research and the general approach that was taken. Chapters II and III concern selection of candidate explanatory variables and initial analysis of these variables. Chapter V details the building of the model and its refinement. The last three Chapters, VI through VIII assess the model's fit, explores a possible model use, and finishes with recommendations and conclusions.

# II. RESEARCH GOALS

## A. APPROACH

USAREC maintains a large historical database containing extensive information on every contract that is signed throughout the Command. The approach of this study was to use this database and other readily available USAREC data resources to develop a DEP attrition model. This approach has resulted in quantitative models that should be useful to USAREC as supplements to field expertise. Research focused on providing the recruiter in the field with a system to complement his subjective opinion as to the risk of a DEP member becoming a loss. Though certain conclusions were drawn regarding USAREC DEP policies, this was not the emphasis.

## B. PREVIOUS RESEARCH EFFORTS

Research was conducted on the DEP loss process during the 1980's. Current USAREC DEP tracking and analysis is aggregated at the Recruiting Battalion level to provide early warning in case accession goals are in jeopardy. Several studies have used time series analysis to predict the rate in which DEP loss occurs [Ref. 5]. A shortcoming with this approach is it assumes DEP losses occur on the date reported in the database. These dates are then used for developing models of DEP loss rates. In actuality, this date merely reflects when the

recruiting chain of command officially reported the loss. The actual date in which the recruit decided to leave the DEP could have been months prior.

Individual contract level models have been developed but focused on only those contracts signed by high school seniors and graduates in the highest mental category.[2] The most recent year of recruiting data used in developing these models was FY 88. Our research used data covering all non prior service contracts signed in FY 86 through FY 90. We examined contributions of the following new areas:

- The 17 - 21 year old population in each Recruiting Battalion's region
- Military/civilian pay ratios for the Recruiting Battalion
- Total number of Department of Defense recruiters in the Recruiting Battalion's region
- Recruiting Battalions
- Career Management Field (CMF) of contract
- Renegotiation status of the contract
- Number of recruiters per contract in the Recruiting Battalion (contract density)
- Brigade (local) and national advertising budgets

The inclusion of these new variables may potentially result in better predicting power as compared to already

---

[2] Nelson, 1988, Army Research Institute and Celeste, 1989, WESTAT.

existing models. Additionally, many officials at USAREC believe the combination of a declining advertising budget, fewer recruiters in the field, and a dwindling 17 - 21 year old population have significantly impacted all recruiting operations over the last five years.[3] All three of these concerns are addressed in the models developed here.

---

[3] This information was obtained during interviews with USAREC personnel from 18 November through 21 December 1990 during an experience tour at USAREC Headquarters, Fort Sheridan, IL.

# III. VARIABLE DEVELOPMENT

There are many similarities between the initial selling of a contract by a recruiter and the reselling that goes on with a member of the DEP. The recruiter must periodically meet with the DEP member and resell him on his initial contract. This recruiting effort receives command emphasis throughout USAREC. For this reason, many of the same variables used in contract production models were analyzed for applicability in a DEP loss model. Explanatory variables can be described as being either individual, demographic, or policy factors.

## A. INDIVIDUAL FACTORS

Individual factors are the personal characteristics of the DEP member. Table I shows the variables that were considered for inclusion and their source. These variables represent the characteristics of the recruit on the day that the contract was signed. USAREC updates the EDUC variable as the DEP member's education status changes. Therefore, this value was obtained from a previous education code in the database. The EDUC variable includes four classes. All education codes indicating education levels above high school were aggregated

Table I INDIVIDUAL FACTORS TO BE ANALYZED

| VARIABLE | DESCRIPTION | SOURCE[1] |
|---|---|---|
| AGE | AGE IN YEARS ON CONTRACT DATE | USAREC MM |
| MARITAL | MARITAL STATUS | USAREC MM |
| SEX | MALE OR FEMALE | USAREC MM |
| RACE | WHITE, BLACK, HISPANIC, ASIAN, OTHER | USAREC MM |
| EDYRS | YEARS OF EDUCATION | USAREC MM |
| EDUC | STATUS OF HIGH SCHOOL DIPLOMA, EITHER IN HIGH SCHOOL, NON GRADUATE, DIPLOMA GRADUATE, OR OTHER TYPE OF GRADUATE | USAREC MM |
| AFQT | ARMED FORCES QUALIFICATION TEST SCORE | USAREC MM |
| CONTDATE | DATE IN WHICH CONTRACT WAS SIGNED | USAREC MM |
| DEPEND | NUMBER OF DEPENDENTS | USAREC MM |

NOTE: 1. USAREC MM is the Minimaster database maintained at USAREC containing information on all contracts signed during a fiscal year.

into one class. Likewise, the many types of high school graduates other than regular diploma graduate were aggregated into one class. RACE was aggregated into the four numerically largest races. The category OTHER included the remaining less populace races.

## B. DEMOGRAPHIC FACTORS

Demographic factors are the characteristics of the geographic region in which the recruit lived when the contract was signed. Table II describes these variables and their sources. Quarterly data were used to calculate these variables. When monthly data were available, as in the MISSION and DOD variables, the quarter's mean was used. The level of

Table II DEMOGRAPHIC VARIABLES TO BE ANALYZED

| VARIABLE | DESCRIPTION | SOURCE[1] |
|---|---|---|
| UNEMP | LOCAL UNEMPLOYMENT RATE IN THE RECRUITING BATTALION IN THE QUARTER IN WHICH THE CONTRACT IS SIGNED | SUPERSITE |
| BN | RECRUITING BATTALION (54 CONSIDERED) | USAREC MM |
| MISSION | RECRUITING BATTALION RATIO: $\frac{\text{MILITARY AVAILABLE 17-21 OLD}}{\text{NUMBER OF CONTRACTS}}$ | USAREC MM / BERLIANT |
| PAYRATE | RECRUITING BATTALION RATIO: $\frac{\text{CIVILIAN MEDIAN INCOME}}{\text{E-2 UNDER 2 YEARS PAY}}$ | SUPERSITE / US ARMY FINANCE |
| DOD | RECRUITING BATTALION RATIO: $\frac{\text{MILITARY AVAILABLE 17-21 OLD}}{\text{MEAN NUMBER OF DOD RECRUITERS}}$ | USAREC PAE |

<u>NOTE:</u> 1. Supersite is the DOD Manpower Data Center's Supersite Demographic Database; USAREC MM is the USAREC Minimaster database; Berliant is an Army Research Institute study [Ref. 6]; USAREC PAE is the USAREC Program Analysis and Evaluation Directorate

the demographic variable is the Recruiting Battalion. PAYRATE was not indexed for inflation. Since civilian median income and E-2 pay increased separately, the ratio of these two incomes was the explanatory variable used. Of the 55 Recruiting Battalions, the San Juan Battalion was eliminated from the study due to lack of demographic data.

The MISSION variable was used to represent contract density in each region. A large value indicates a high output Recruiting Battalion relative to their available population base. It also might indicate a propensity of candidates in the region to join the US Army.

The DOD variable was included to allow for the presence of Department of Defense recruiters. Small values in this variable would represent competition from the other services for the available recruit population. Many USAREC officials postulate that there is an increased propensity to join the US Army when any service is well represented in a region.

## C. POLICY FACTORS

Policy factors are those characteristics of the contract that are dependent on USAREC policies current at the time the contract was signed. Table III describes these factors and their sources. Note that the TIMEDEP variable is the contracted time to be in the DEP, not the actual time. As with

**Table III** POLICY VARIABLES TO BE ANALYZED

| VARIABLE | DESCRIPTION | SOURCE[1] |
|---|---|---|
| TIMEDEP | TIME CONTRACTED TO BE IN THE DEP | USAREC MM |
| BONUSAMT | AMOUNT OF BONUS ( IF ANY ) | USAREC MM |
| RENO | BINARY VARIABLE INDICATING IF A CONTRACT RENEGOTIATION OCCURRED WHILE IN THE DEP | USAREC MM |
| ACF | INDICATES IF THE RECRUIT IS AN ARMY COLLEGE FUND TAKER | USAREC MM |
| CMF | CAREER MANAGEMENT FIELD (31 AVAILABLE) | USAREC MM |
| TERM | TERM OF CONTRACTED ENLISTMENT | USAREC MM |
| CONPER | CONTRACTS PER RECRUITER FOR THE QUARTER IN THE RECRUITING BATTALION | USAREC PAE |
| BDEADV | BRIGADE LOCAL ADVERTISING BUDGET FOR THE FISCAL YEAR AND RECRUITING BRIGADE | USAREC APAD |
| NATADV | NATIONAL ADVERTISING BUDGET FOR FISCAL YEAR | USAREC APAD |

NOTE: 1. USAREC MM is the Minimaster database; USAREC PAE is USAREC Program Analysis and Evaluation Directorate; USAREC APAD is USAREC Advertising and Public Affairs Directorate.

demographic factors, CONPER is the quarterly mean with respect to both number of contracts and the number of recruiters. Data were aggregated at the Recruiting Battalion level. The BDEADV and NATADV advertising variables were indexed to FY 86 dollars using USAREC Advertising and Public Affairs Directorate advertising price indexes.

## D. DATABASE

### 1. Sources

As shown in Tables I through III, the USAREC Minimaster database was the primary source of data for this model development. These records are year end pictures of all recruiting contract activity during the fiscal year. Contracts are represented on successive fiscal year Minimaster files until the contract is closed by either accession or DEP loss. An example: a contract signed in FY 86 with an accession or DEP loss in FY 87 would be on both Minimaster 86 and 87 databases. Minimaster 86 would indicate this as an open record. Then, Minimaster 87 would contain the accession status of the contract.

Minimaster 86 did not include the bonus amount of the contract but only whether one was received. Using historical bonus information from USAREC Recruiting Operations Directorate, these data were reconstructed.

Information regarding US Army and DOD recruiter field strength and advertising budgets was obtained from

directorates at USAREC Headquarters. DOD Manpower Data Center (DMDC) provided the employment and civilian median income information for each Recruiting Battalion. DMDC subcontracted to provide USAREC with a Supersite system which aggregates county level economic data to Recruiting Battalion level [Ref. 7]. The source for the 17 - 21 year old prime recruiting market at the battalion level was a 1989 Army Research Institute study conducted by Kenneth R. Berliant [Ref. 6].

2. Database Development

Statistical Package for Social Scientists (SPSS) was used for screening, sorting, and merging the Minimaster records in preparation for model development. This statistical package was used because of its widespread use at USAREC. This should assist any future updating of the model as data become available. Table IV details the results of the database after screening for unwanted records and data errors. A total of 247,592 records were eliminated as being open, prior service, from the San Juan Battalion, or contracts signed before FY 86. Open records were not closed out in the given fiscal year as a result of accession or DEP loss. They were then repeated and closed out in the following fiscal year. Approximately 3.5% of the records were eliminated due to coding errors in the data. Due to the large size of the database, 715,668 records, it was not felt that this would significantly bias the data or the analysis results. Analyses indicated that the eliminated

records possessed approximately the same percentage of DEP losses as the entire contract population.

After the Minimaster files were screened and concatenated, the demographic and policy variables containing quarterly values were merged to create the final large database. There were 689,278 contract records available, each containing DEP loss status and values of 24 candidate explanatory variables.

Table IV RESULTS OF DATABASE SCREENING

| RECORDS INITIALLY AVAILABLE | NUMBER |
|---|---|
| MINIMASTER FY86 .......................... | 208,504 |
| MINIMASTER FY87 .......................... | 206,326 |
| MINIMASTER FY88 .......................... | 192,048 |
| MINIMASTER FY89 .......................... | 193,682 |
| MINIMASTER FY90 .......................... | 162,700 |
| SUBTOTAL | 963,260 |
| **RECORDS ELIMINATED** | |
| OPEN RECORDS [1]................................ | 112,293 |
| PRIOR SERVICE RECORDS ........................ | 66,201 |
| CONTRACTS SIGNED IN FY85 ..................... | 60,680 |
| RECORDS FROM SAN JUAN BATTALION .............. | 8,418 |
| SUBTOTAL | 247,592 |
| **RECORDS ELIMINATED DUE TO ERRORS IN DATA** | |
| NUMBER OF DEPENDENT ERRORS ................... | 12,467 |
| BATTALION / BRIGADE DESIGNATION ERRORS ....... | 4,846 |
| TERM OF SERVICE ERRORS ....................... | 2,195 |
| NUMBER OF YEARS EDUCATION ERRORS ............ | 1,907 |
| CONTRACT YEAR / MONTH ERRORS ................ | 1947 |
| PROJECTED ACCESSION YEAR / MONTH ERRORS ...... | 1716 |
| BIRTH YEAR / MONTH ERRORS ................... | 579 |
| TIME IN DEP ERRORS .......................... | 512 |
| MILITARY OCCUPATION SPECIALTY ERRORS ......... | 130 |
| ARMED FORCES QUALIFICATION TEST ERRORS ....... | 91 |
| SUBTOTAL | 26,390 |
| | |
| **RECORDS AVAILABLE FOR ANALYSIS TOTAL** | 689,278 |

NOTES: 1. Open records have not been closed out in the given fiscal year as a result of accession or DEP loss. They are then repeated and closed out in the following fiscal year.

# IV. DATA SUMMARY

## A. DEP LOSS TRENDS

An initial analysis with data in the DEP loss database concerned possible seasonal effects on DEP losses during the Recruiting year. Two methods were used to calculate the DEP loss percentages. The first method, shown in Figure 2, was by contract cohort. Contracts for the months of FY 86 through FY 90 were tracked as a cohort. Percent DEP loss is the

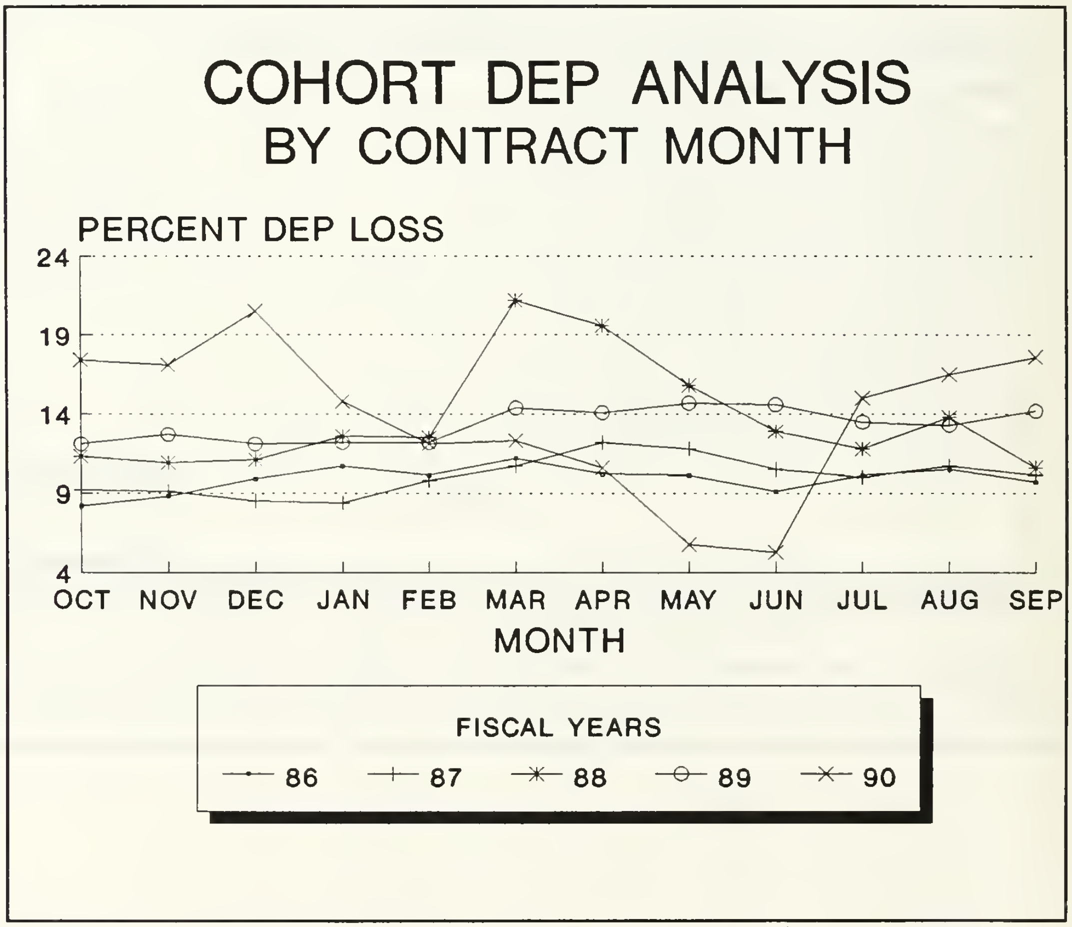


Figure 2 Contract Cohort DEP Loss Analysis

percentage of this cohort that resulted in a DEP loss. There did not appear to be any strong reoccurring seasonal trend. The significant increase in DEP loss in the spring of 1988 was a result of a one time DEP forgiveness program instituted by USAREC in response to accession cutbacks.

The second method for examining DEP loss was by accession cohort. The accession status of all recruits that were projected to access in the months of FY 86 through FY 90 were tracked. The percent of the accession cohorts that resulted in DEP loss is depicted in Figure 3.

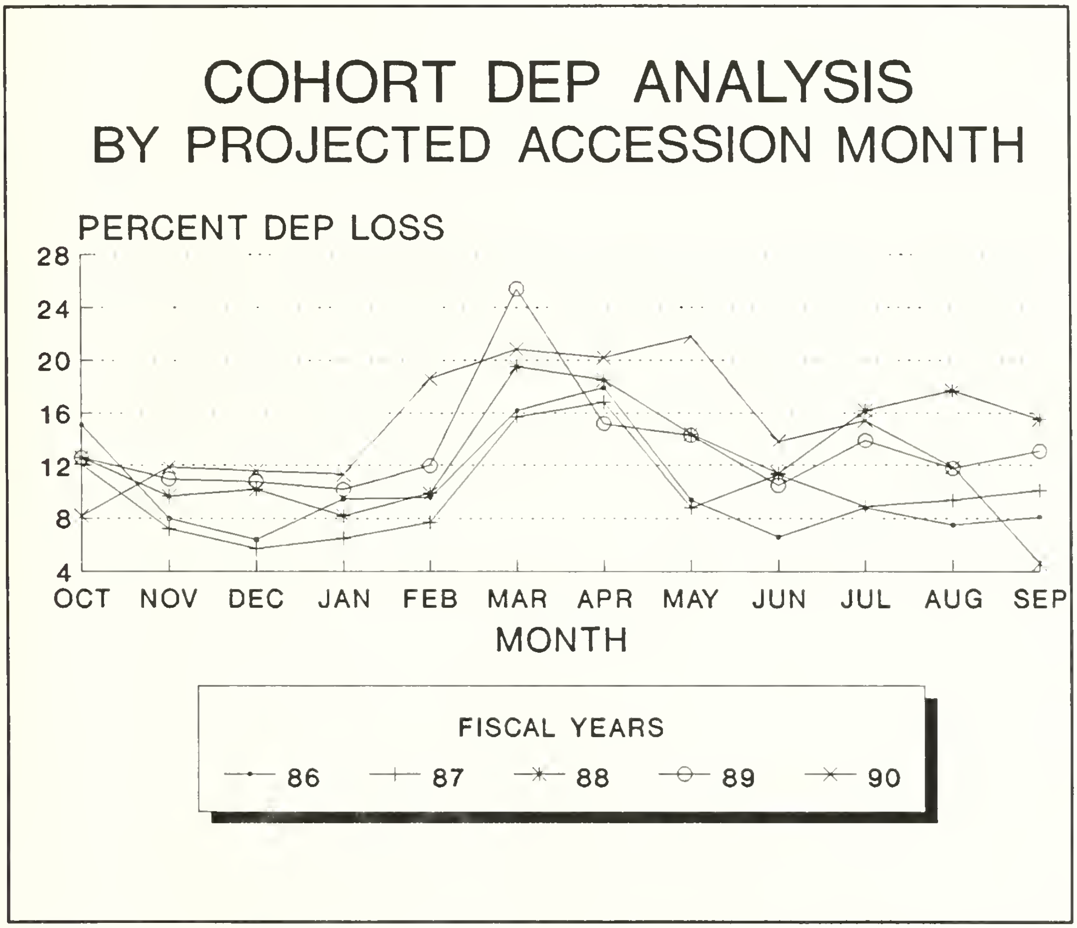


Figure 3 Projected Accession Cohort DEP Loss Analysis

There appeared to be a trend for higher DEP losses in spring, March through May, during each of the five fiscal years. This may have been a result of high school seniors who signed contracts early in the year. They then may have changed either education or career goals in the spring. Since there appeared to be a seasonal trend, a dummy variable for projected accession month was included in the model development.

## B. INTERVAL VARIABLES

Fourteen of the 23 initial explanatory variables were interval (scale) variables. Using SPSS, initial analyses were conducted to determine if there were significant differences between the two groups, accession and DEP loss, with respect to these variables. The mean values for the two groups are listed in Table V. The T-test is used as a basis for rejecting or failing to reject the null hypothesis that the two sample means are equal. Due to the large sample size (689,278), the T-test does not require that the samples come from a Normal population. With T-test significance levels below .00005 for these interval variables, there is less than .005% chance that such sample means would be this different if the population means were equal. We acknowledge that with this large sample that the null hypothesis will almost always be rejected. Though statistical significance is indicated, we believe there is practical significance in the difference of these means.

Table V INTERVAL VARIABLE ANALYSIS

| INTERVAL VARIABLE | VARIABLE DESCRIPTION | ACCESSION CONTRACTS [3] | DEP LOSSES [3] |
|---|---|---|---|
| AGE | AGE IN YEARS ON CONTRACT DATE | 19.9572 | 19.7859 |
| EDYRS | YEARS OF EDUCATION | 12.0702 | 11.6019 |
| AFQT | ARMED FORCES QUALIFICATION TEST PERCENTILE SCORE | 58.6334 | 59.7147 |
| TERM | TWO THROUGH SIX YEARS OF CONTRACTED SERVICE | 3.539 | 3.5922 |
| BONUSAMT | CONTRACT BONUS AMOUNT (IF ANY) | 318.97 | 283.27 |
| DEPEND | NUMBER OF DEPENDENTS | .1782 | .0820 |
| TIMEDEP | TIME CONTRACTED TO BE IN THE DEP | 3.973 | 5.898 |
| UNEMP | LOCAL (BN) UNEMPLOYMENT RATE AT TIME OF CONTRACT | 6.355 | 6.06 |
| MISSION [1] | RATIO: MILITARY AVAIL 17-21 YEAR OLD (BN) / NUMBER OF CONTRACTS (BN) | 394.65 | 412.83 |
| PAYRATE [1] | RATIO: CIVILIAN MEDIAN INCOME (BN AREA) / MILITARY PAY (E-2 UNDER 2 YEARS) | 2.872 | 2.937 |
| COMPER [1] | RATIO: NUMBER OF CONTRACTS (BN) / MEAN # OF RECRUITERS ASSIGNED (BN) | 8.24 | 7.58 |
| DOD [1] | RATIO: MILITARY AVAIL 17-21 YEAR OLD (BN) / MEAN # OF DOD RECRUITERS (BN) | 767.85 | 760.3 |
| BDEADV [2] | BRIGADE LOCAL ADVERTISING BUDGET FOR THE FISCAL YEAR | 890,607 | 872,658 |
| NATADV [2] | USAREC NATIONAL ADVERTISING BUDGET FOR FISCAL YEAR | 65,093,198 | 63,654,535 |

NOTES: 1. Variables are calculated using data for quarter in which contract was signed. 2. Variables are calculated for fiscal year in which contract was signed. 3. T-test significance less than .00005

The variable TERM is the only variable in which the practical significance appears questionable.

The mean values for these interval variables give some insight into the DEP loss contract holder, compared to those who access. The DEP loss is slightly younger and has fewer

years of education because he may be more likely to still be in high school. His AFQT score is higher than average contracts which may indicate more opportunities. His contract term of service is longer and he gets less than an average bonus amount. He has fewer dependents to worry about and is planning on spending much more than average time in the DEP awaiting accession onto active duty. The economic situation in his Recruiting Battalion region is better than average as indicated by lower unemployment and better civilian pay. There is less contract density in his Recruiting Battalion region. There are more DOD recruiters in his region than average. USAREC spends less on advertising in his region of the country.

The CONPER values appeared counter intuitive. The number of contracts per recruiter was lower for DEP loss contract holders. This may indicate that high mission recruiters tended to have less DEP losses. This phenomena may be due to USAREC's Recruiting Zone Analysis (RZA) that assigns recruiters and missions to Recruiting Battalions. This could indicate that high propensity regions as determined by RZA suffer less DEP losses.

As previously mentioned, the large database assisted in increasing the significance of these T-tests. This may have overemphasized their explanatory value as covariates in attrition models. Even so, these interval variables appeared

significant in the univariate analyses and were included as candidate explanatory variables in the modeling process.

## C. CLASS VARIABLES

The remaining nine explanatory variables were categorical or class variables. Again, using SPSS, cross tabulations with Chi-Square tests were conducted to determine if DEP loss status was independent of the class variables. Table VI lists the first seven class variables and Appendix A, Tables XIII through XVI list the class variables with larger numbers of levels, Career Management Field (CMF) and Recruiting Battalion. The results of the Chi-Square tests indicated that all the class variables were highly significant. As with the interval variables, there is less than a .005% chance that such distributions would have occurred if DEP loss status was independent of these class variables.

Initial analyses indicated that marital status, sex, education level, and contract renegotiation status were the more significant explanatory class variables. Several of the CMF's and Recruiting Battalions appeared to be strong explanatory variables. CMF 00 had a 99.4% DEP loss rate. According to USAREC Recruiting Operations Directorate, this is not a valid CMF. It was used in FY 87 and FY 88 as a surrogate CMF for known DEP losses who were not officially dropped for an extended period. This use of CMF 00 freed the previously

Table VI CLASS VARIABLE ANALYSIS

| CLASS VARIABLE | VARIABLE DESCRIPTION | PERCENT ACCESSION[1] | PERCENT DEP LOSS[1] |
|---|---|---|---|
| MARITAL | MARITAL STATUS TIME OF CONTRACT | | |
| ..MARRIED | 9.6% MARRIED | 10.3 | 4.56 |
| ..SINGLE | 90.4% SINGLE / NOT MARRIED | 89.7 | 95.44 |
| SEX | MALE OR FEMALE | | |
| ..MALE | 84.6% MALE | 85.5 | 78.24 |
| ..FEMALE | 15.4% FEMALE | 14.5 | 21.76 |
| RACE | FOUR LARGEST RACES AND OTHER | | |
| ..WHITE | 70.1 % WHITE | 69.6 | 73.6 |
| ..BLACK | 24.4% BLACK | 24.9 | 20.8 |
| ..HISPAN | 2.3% HISPANIC | 2.34 | 2.15 |
| ..ASIAN | 1.2% ASIAN | 1.2 | 1.05 |
| ..OTHER | 2.0% OTHER / UNKNOWN | 1.96 | 2.4 |
| EDUC | EDUCATION CODE AT CONTRACT | | |
| ..SENIOR | 29.7% IN SCHOOL | 27.6 | 45.66 |
| ..NONGRAD | 3.8% NON-GRADUATE HIGH SCHOOL | 3.86 | 2.84 |
| ..DIPGRAD | 62.5% DIPLOMA GRAD HIGH SCHOOL | 64.4 | 48.1 |
| ..OTHGRAD | 4.1% OTHER TYPE GRAD HIGH SCHOOL | 4.14 | 3.4 |
| ACF | ARMY COLLEGE FUND TAKER | | |
| ..TAKER | 18.9% ACF TAKERS | 19.04 | 17.63 |
| ..NOTAKER | 81.1% NOT ACF TAKERS | 80.96 | 82.37 |
| RENO | RENEGOTIATION OF CONTRACT IN DEP | | |
| ..YESRENO | 8.9% OF CONTRACTS RENEGOTIATED | 8.22 | 13.95 |
| ..NORENO | 91.1% NOT RENEGOTIATED | 91.78 | 86.05 |
| RECFY | RECRUITING FISCAL YEAR IN WHICH CONTRACT WAS SIGNED | | |
| ..86 | 26.5% SIGNED IN FY86 | 27.09 | 21.76 |
| ..87 | 24.1% SIGNED IN FY87 | 24.49 | 21.54 |
| ..88 | 19.0% SIGNED IN FY88 | 18.64 | 21.76 |
| ..89 | 20.3% SIGNED IN FY89 | 20.05 | 22.13 |
| ..90 | 10.1% SIGNED IN FY90 | 9.73 | 12.83 |
| TOTAL[2] | TOTAL CONTRACT PERCENTAGES | 88.17 | 11.83 |

NOTES: 1. Cell difference significance less than .00005 Chi-square test. 2. Class variable analysis for Career Management Field (CMF) and Battalions see Appendix A, Tables XIII through XVI.

reserved CMF to be used for another contract. Rather than delete these records and loose the data, they were retained and dealt with during model development.

The results of the data assessment process justified inclusion of the 23 candidate explanatory variables. It also revealed that due to a seasonal trend, the projected accession month may be a strong explanatory variable. In our model development we attempted to use these 24 interval and categorical variables to predict DEP loss.

# V. MODEL DEVELOPMENT

## A. MODEL SELECTION

Empirically, the individual process of attrition from the DEP is represented by a dichotomous (binary) dependent variable which categorizes individuals either as accessions or DEP losses. The dependent variable definition is as follows:

$$Y_i = \begin{cases} 0 \text{ , if individual i accesses into the US Army} \\ 1 \text{ , if individual i is a DEP loss.} \end{cases}$$

Logit models are particularly well suited for dichotomous dependent variables because the logistic distribution lends itself to a meaningful interpretation. For notational purposes, the quantity:

$$\pi(X) = E( Y \mid X ) \qquad (1)$$

is used to represent the conditional mean of Y (DEP loss or accession) given the covariates X (explanatory variables).

The specific form of logistic regression model we used is as follows:

$$\pi(X) = E(Y|X) = \frac{1}{1 + e^{-g(X)}} \qquad (2)$$

where g(X) is the linear combination:

$$g(X) = B_0 + B_1x_1 + B_2x_2 + \ldots + B_px_p \qquad (3)$$

Where p is the number of covariates, $x_i$ $i=1,\ldots,p$ are the covariates, $X = (x_1, x_2, \ldots, x_p)$, $B_0$ is the constant parameter, and $B_i$ $i=1,\ldots,p$ are the coefficient parameters.

The conditional mean in equation **(1)** is bounded in value by zero and one because of the fraction on the right hand side of equation **(2)**. The usefulness of logistic regression is that the value, $\pi(X)$ may be interpreted as the probability of being a DEP loss (Y=1) given explanatory variables X, or $P(Y=1|X)$.

The logit transformation used in the fitting of the model is:

$$g(X) = \ln\left[\frac{\pi(X)}{1 - \pi(X)}\right] \qquad (4)$$

This logit, g(X) is linear in its parameters, is a continuous variable ranging in value from negative infinity to infinity. In order to estimate the value of $\pi(X)$ the parameters $B_0$ through $B_p$ from equation **(3)** must be estimated using the method of maximum likelihood. [Ref. 8:p. 1-11]

The method of maximum likelihood uses the known covariates, X, to compute the estimates for $B_0$ through $B_p$ so as to maximize the likelihood of obtaining the observed DEP loss status (Y=0 or 1). For a sample of size n, let $y_i$ and

$X_i = (x_{1i}, x_{2i}, \ldots, x_{pi})$ be the observed DEP loss status and vector of corresponding covariates for individual i, i=1,...,n. The likelihood (normal) equation resulting from the method of maximum likelihood for $B_0$ is :

$$\sum_{i=1}^{n} [y_i - \pi(X_i)] = 0 \tag{5}$$

Similarly, the normal equations for $B_1$ through $B_p$ are:

$$\sum_{i=1}^{n} x_{ji} [y_i - \pi(X_i)] = 0 \quad ; j = 1,2,\ldots,p \tag{6}$$

The value of the vector $B=(B_0, B_1, \ldots, B_p)$ given by the solution of these p+1 equations is $\hat{B}$ , the maximum likelihood estimator for B. The values for the estimated probability of DEP loss are obtained from equations (2) and (3) by replacing B with $\hat{B}$. The estimated probability of DEP loss is denoted $\hat{\pi}$. An interesting result of equation (5) is the following:

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{\pi}(X_i) \tag{7}$$

The sum of the n observed values, $y_i$, is equal to the sum of the n predicted (expected) values, $\hat{\pi}_i$. This property of logistic regression was exploited in our assessment of the fit of the model. The solution of the normal equations above is found by an iterative process which has been programmed into

many available logistic regression computer software packages such as SPSS. The development and rationale for this model is given in Reference 8, pages 8-11.

## B. MODEL BUILDING

SPSS, version 4.0, Logistic Regression Procedure was used to fit the model. This procedure required recoding of the class (categorical) variables. The following class variables with two levels were recoded (0,1) to indicate the presence of an attribute: MARITAL (married=1), SEX (female=1), ACF (yes=1), and RENO (yes=1). The other six class variables were recoded using the deviation coding scheme [Ref. 9:p. 55]. The number of new dummy variables required to represent a class variable with n levels is n-1. For the deviation coding scheme, if any of first n-1 levels of a class variable were present its corresponding new dummy variable was assigned the value of one. Otherwise, the new dummy variable was assigned the value of zero. In order to represent the presence of the nth level of a class variable, all the n-1 new dummy variables were assigned the value of negative one. This resulted in the creation of 105 new variables to represent RACE, EDUC, RECFY, BN, CMF, and PADDMO.

### 1. Variable Selection

SPSS's Logistic Regression procedure has the capability of executing stepwise variable selection. We used the forward stepwise selection as a basis for building our

model. The algorithm commenced with only the constant term in the model. Then, the variable with the lowest significance level for the Score statistic, provided it was lower than the chosen cutoff value $P_{in}$, was entered into the model. The Wald statistic's significance level was used to examine variables for possible elimination [Ref. 9:p. 56]. If the Wald statistic's significance level was higher than $P_{out}$, the variable was eliminated from the model. If no variable met the elimination criteria, the next eligible variable was added. This process continued until either a previously selected model was encountered or there were no further variables meeting the entry or removal criteria. Dummy variables representing the different levels of a class variable entered or were removed from the model as a group. [Ref. 9:p. 56-57]

Hosmer and Lemeshow [Ref. 8:p. 88] suggest the use of $P_{in} = .15$ and $P_{out} = .20$ as the best criteria for use in stepwise logistic regression using the Wald statistic. These criteria were aimed at selection of important variables for the model while also providing a parsimonious model.

Due to the computationally intensive nature of the iterative algorithms used to fit the model, combined with the numerous models built in forward stepwise regression, only a random 10% sample (68,962 cases) of the database was used in variable selection. This sample size required nearly 24 hours of CPU time on an Amdahl 5990-500 mainframe computer.

Variable selection resulted in all variables meeting the $P_{in}$ / $P_{out}$ criteria except two: these variables, MISSION and RECFY, were excluded from the model. The MISSION (contract density) variable's exclusion may have been a result of Recruiting Zone Analysis (RZA) used in assigning contract quotas to the Recruiting Battalion. RZA uses many of the same explanatory variables as our fitted model to determine each Recruiting Battalion's contract density. Therefore, this MISSION variable may not have provided the fitted model with information not already supplied by other explanatory variables. The non-selection of RECFY (Recruiting Fiscal Year) by the stepwise procedure may indicate that there was not a strong yearly influence on DEP loss that was not represented by one of the other chosen explanatory variables. This exclusion could prove to be helpful in future prediction uses of the model.

### 2. Interaction Terms

Univariate analyses and insight into the recruiting environment suggested that consideration of certain interaction terms was appropriate. A dozen interaction terms including combinations of RACE, EDUC, DEPEND, SEX, and MARITAL were considered. Only the RACE by EDUC interaction term was significant with respect to $P_{in}$ in the stepwise procedure. The

inclusion of this interaction term did not result in the removal or entry of any previously selected or non-selected variables.

### 3. Scaling

The continuous scaled interval variables were checked for the assumption of linearity in the logit, g(X), in equation (3). To this point all the interval variables, less MISSION, were identified as significant. Scaling assisted in obtaining the correct parametric relationship during the model refinement stage. We used the Box-Tidwell transformation to evaluate the need for scaling [Ref. 8:p. 90]. This simple technique adds a term of the form $x \cdot \ln(x)$ to the model for each continuous scaled interval variable. If the coefficient of these new variables appeared significant, there was evidence of non-linearity in the logit.

This technique resulted in six of the thirteen selected class variables, EDYRS, TIMEDEP, AGE, UNEMP, CONPER, and DOD indicating possible non-linearity. This technique could not be used for BONUSAMT and DEPEND because they included many values of zero. Therefore, these two variables were also included for further analysis.

A technique proposed by Hosmer and Lemeshow [Ref. 8:p. 90] was used in identifying the need to introduce new, higher-order variables in the model as a scaling method for those variables indicating possible non-linearity. The range of each

of these independent continuous interval variable was broken into groups and treated as a class (categorical) variable. Each case was assigned to the categorical class that represented its range in the original interval scale. The group representing the lowest scaled values served as the referent group. A model was fit to the same 10% random sample of the database using univariate logistic regression with only the one categorical variable. We then plotted the estimated coefficients for the levels of the categorical variable versus the group midpoint values from the initial interval scale. We chose the most logical shape for the scaling of the independent variable.

Figure 4 illustrates the results of using this technique on EDYRS (years of education). The unusual shape of the curve suggested that those in the DEP with eleven years of education had a higher probability of becoming a DEP loss. Likewise, DEP members with substantially more or less than eleven years of education appear to be at a greater risk of DEP loss relative to those with only several years more or less than eleven years of education.

We created a new variable, EDYRS2, representing |EDYRS-11| , the distance from eleven years education. Model log-likelihood, covered in more detail in Chapter 6, was used to compare the improvement of introducing new higher order terms. The larger the model log-likelihood statistic, the more likely that if the fitted model is the correct model the

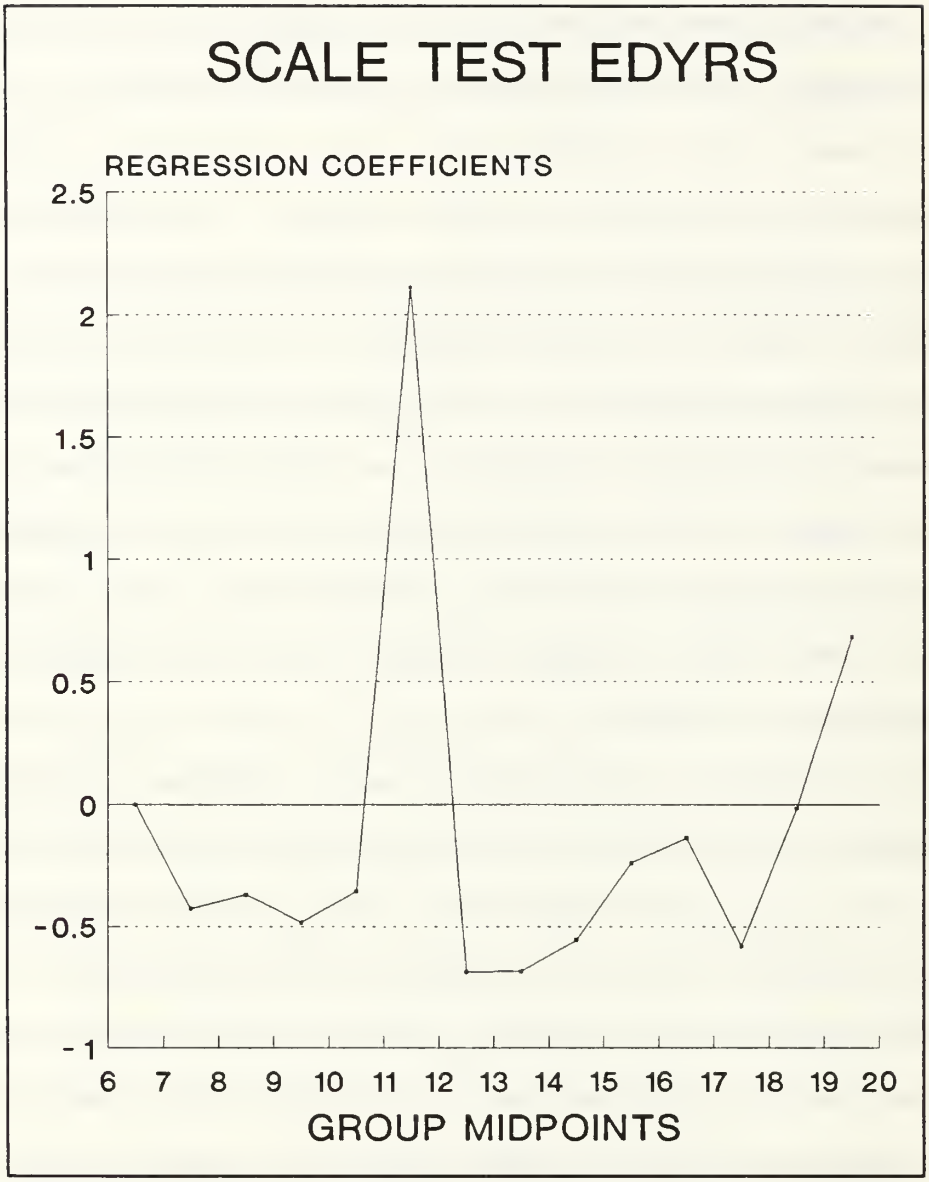

Figure 4 Hosmer-Lemeshow Scale Analysis on EDYRS

observed results would be obtained given the estimated parameters, $\hat{B}$. Univariate analysis indicated that EDYRS2 alone more than doubled the model log-likelihood over EDYRS by itself.

The same Hosmer-Lemeshow grouping technique was used for EDYRS2 to determine the need for introduction of higher order terms. Figure 5 depicts this new assessment. This curve appeared to be quadratic in the logit. A quadratic term, EDYRS22 = $(\mathrm{EDYRS2})^2$ was added to the model. The model containing EDYRS2 and EDYRS22 doubled the model log-likelihood again and was more than four times larger than the model containing EDYRS alone. Similar analyses were conducted on the

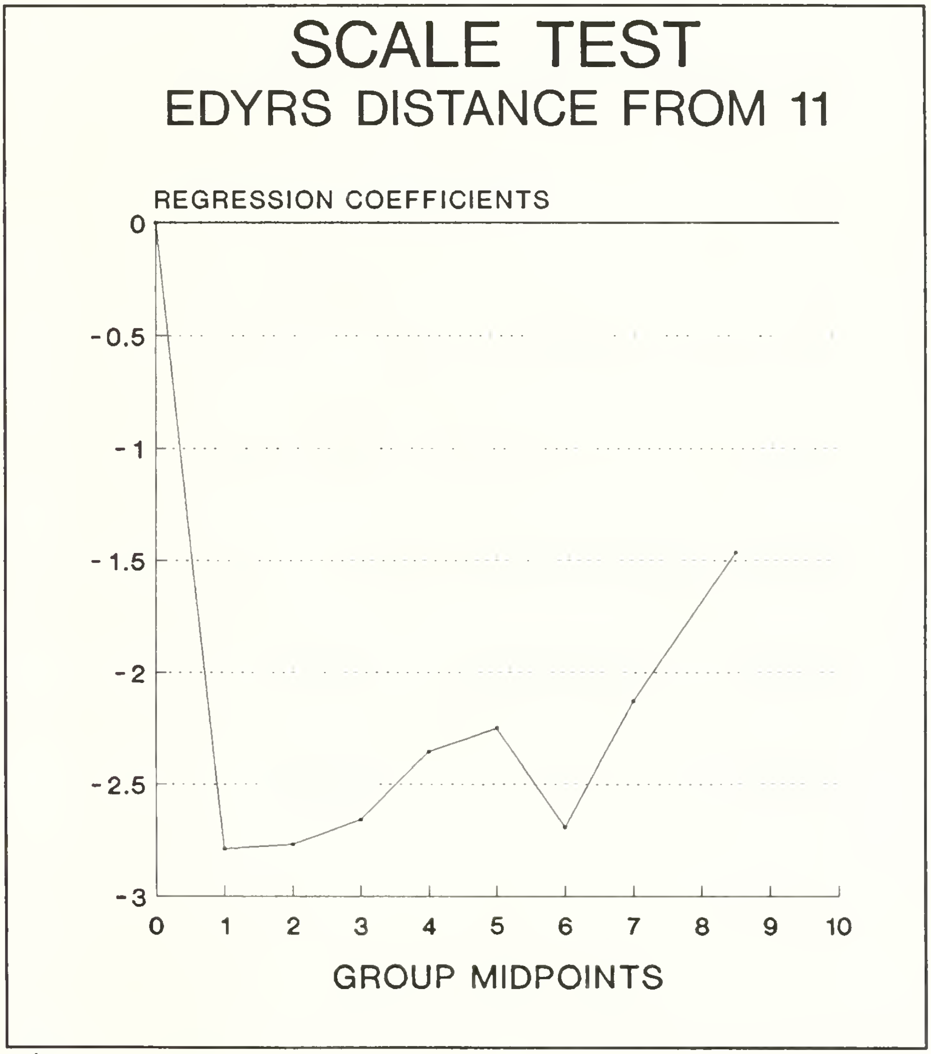


**Figure 5** Hosmer-Lemeshow Scale Analysis on EDYRS2

other seven continuous variables for which non-linearity in the logit was suspected. Five of these seven assessments resulted in the scaling depicted in Table VII.

As a result of the addition of these higher-order terms, the variables BONUSAMT, DOD, DOD2, and DOD3 were eliminated from the model using backward stepwise elimination. The same values $P_{in} = .15$, $P_{out} = .20$ as in forward stepwise selection were used.

Table VII RESULTS OF HOSMER-LEMESHOW SCALING

| ORIGINAL VARIABLE | SCALING | NEW VARIABLES | IMPROVEMENT RESULTS[1] |
|---|---|---|---|
| TIMEDEP | CUBIC | $TIMEDEP2 = (TIMEDEP)^2$<br>$TIMEDEP3 = (TIMEDEP)^3$ | 3.6 % |
| AGE | CUBIC | $AGE2 = (AGE)^2$<br>$AGE3 = (AGE)^3$ | 1690 % |
| DEPEND | CUBIC | $DEPEND2 = (DEPEND)^2$<br>$DEPEND3 = (DEPEND)^3$ | 31.3 % |
| CONPER | QUADRATIC | $CONPER2 = (CONPER)^2$ | 28.8 % |
| DOD | CUBIC | $DOD2 = (DOD)^2$<br>$DOD3 = (DOD)^3$ | 445 % |

NOTES: 1. Improvement is the percent increase in the model log-likelihood of the fitted model containing the new higher-order variables over a fitted model containing only the original variable.

## C. MODEL EXECUTION

The final DEP loss model contained 23 interval scaled variables, five categorical (class) variables represented by 101 dummy variables, one interaction term with 12 levels, and the constant term. The total number of coefficients estimated, components of the B vector, was 136. Table VIII and Appendix A, Tables XVII through XX contain the variables in the final model, their estimated coefficients, $\hat{B}_i$ and their significance levels based on the Wald statistic. A 25% sample (170,685 cases) was used for estimating the final model's coefficients. Estimation of B with this sample size required the maximum available scratch workspace and almost 20 hours of CPU time on a Amdahl 5990-500 mainframe computer.

Table VIII RESULTS OF FINAL MODEL

| VARIABLE | ESTIMATED COEFFICIENT $\hat{B}_i$ | SIGNIFICANCE LEVEL |
|---|---|---|
| TIMEDEP | .4795 | .0000 |
| TIMEDEP2 | -.0193 | .0000 |
| TIMEDEP3 | 1.7E-05 | .955 |
| AGE | .7221 | .0067 |
| AGE2 | -.0137 | .2147 |
| AGE3 | 4.1E-05 | .7842 |
| EDYRS2 | -4.6175 | .0000 |
| EDYRS22 | .7342 | .0000 |
| SEX | .6336 | .0000 |
| NATADV | -5.5E-08 | .0000 |
| MARITAL | -.398 | .0000 |
| CONPER | -.0335 | .0000 |
| CONPER2 | .001 | .1036 |
| DEPEND | -1.5442 | .0000 |
| DEPEND2 | .9147 | .0000 |
| DEPEND3 | -.1617 | .0000 |
| RENO | -.2474 | .0000 |
| UNEMP | -.0415 | .001 |
| AFQT | .0013 | .0386 |
| BDEADV | 1.3E-07 | .1065 |
| PAYRATE | -1.3486 | .0000 |
| AGE | -.038 | .1836 |
| TERM | -.0426 | .0022 |
| EDUC | NOTE 1 | .0000 |
| RACE | NOTE 1 | .0000 |
| PADDMO | NOTE 1 | .0000 |
| BN | NOTE 1 | .0000 |
| CMF | NOTE 1 | .0000 |
| RACE by EDUC | NOTE 1 | .0002 |
| CONSTANT | -3.691 | .1311 |

NOTES: 1. The estimated coefficients for these class variables were not presented in this table due to their large number of levels. They are located in Appendix A, Tables XVII through XX.

# VI. ASSESSING MODEL FIT

A known problem with the use of logistic regression models is the difficulty in assessing the fit of the computed model. Concerning logistic regression, Dr. Steven Fienberg, says, "But as long as some of the predictors are not categorical, we cannot carry out an omnibus goodness-of-fit test for a model." [Ref. 10:p. 104]. Our fitted model contains 23 interval (non-categorical) variables. Even though we acknowledge this stated difficulty, we attempted to use several known methods to access the fit of our model. We pursued this effort in the hopes of gaining insight into our model's strengths and weaknesses.

## A. LOG-LIKELIHOOD

The SPSS software uses the log-likelihood method to assess the quality of fit of the logistic regression model. With this method, one determines the likelihood of the observed results as a function of the parameter estimates. Since this likelihood is a small value, between zero and one, -2 times the log of the likelihood is used (-2LL). Additionally, the reason -2LL is used is that it is asymptotically Chi-Square distributed. A good model results in a high likelihood or, equivalently, a small value for -2LL. [Ref. 9:p. 52]

Under the null hypothesis that our theoretical model fits perfectly, the value -2LL is from a Chi-Square distribution

with N - p = 170,548 degrees of freedom. Here, N is the number of cases in our 25% sample (170,685) and p is the number of parameters estimated (137). The log-likelihood assessment output from SPSS is depicted in Table IX.

Table IX MODEL LOG-LIKELIHOOD FROM SPSS

| | CHI - SQUARE | DEGREES OF FREEDOM | SIGNIFICANCE |
|---|---|---|---|
| - 2LL | 85,421.7 | 170,548 | .0000 |
| MODEL CHI - SQUARE | 35,812.5 | 137 | .0000 |

The extremely small significance level for - 2LL indicates our model is not a perfect model. The probability that such results would be obtained with the correct model is nearly zero. The model Chi-Square is used to test the null hypothesis that the coefficients of all the variables in the model are zero. The small significance level computed for the model Chi-Square indicates that not all of these coefficients are zero. As noted in the T-tests of Chapter IV, we acknowledge that since the sample size is so large, the null hypothesis that the coefficients are zero will almost always be rejected. Though the null hypothesis of perfect fit of the model was rejected, the null hypothesis that the coefficients are all zero was also rejected.

## B. PEARSON CHI-SQUARE

Hosmer and Lemeshow [Ref. 8:p. 140-145] developed a method for assessing the fit of logistic regression models using a test statistic similar to the Pearson Chi-Square test statistic. The strategy entails grouping the cases by their estimated probabilities, $\hat{\pi}$. Due to our large sample size, we used 20 groups with approximately 8,543 cases per group. The first group contained the 8,543 smallest $\hat{\pi}$ values, the second group the next largest 8,543 values, and so on.

For the y=1 row, representing all contracts that resulted in DEP loss, the expected number of DEP loss contracts for each of the 20 groups was obtained by summing the estimated probabilities of DEP loss, $\hat{\pi}$ for all the members of each of the corresponding 20 groups. The observed values for each of the 20 groups in this row are the number of observed DEP loss contracts within the respective group ($y_i$=1).

With the y=0 row, representing all contracts that resulted in accession, the expected number of accessions for each of the 20 groups was obtained by summing one minus the estimated probability of DEP loss, $\hat{\pi}$ for all the members of each of the corresponding 20 groups. The observed values for each of the 20 groups in this row are the number of observed contracts that resulted in accession within the respective group ($y_i$=0). Table X displays the results of these calculations.

Table X HOSMER-LEMESHOW GOODNESS OF FIT TABLE

| | | LEVELS OF RISK | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 1 | 8 | 5 |
| | # OF CASES | 8545 | 8543 | 8543 | 8543 | 8543 |
| ACCESSION STATUS | $\hat{\pi}(x)$ CUTOFF | .0045 | .0112 | .0173 | .023 | .028 |
| $Y_i = 1$ DEP LOSS | OBSERVED | 379 | 356 | 222 | 235 | 215 |
| | EXPECTED | 17.5 | 66.7 | 122.1 | 172.5 | 219.7 |
| | TEST STAT | 7463 | 1264 | 83 | 23.1 | .103 |
| $Y_i = 0$ ACCESSION | OBSERVED | 8166 | 8166 | 8321 | 8308 | 8328 |
| | EXPECTED | 8527 | 8477 | 8421 | 8370 | 8323 |
| | TEST STAT | 15.3 | 9.9 | 1.2 | .48 | .003 |

| | | LEVELS OF RISK | | | | |
|---|---|---|---|---|---|---|
| | | 8 | 8 | 8 | 8 | 10 |
| | # OF CASES | 8543 | 8543 | 8543 | 8543 | 8543 |
| ACCESSION STATUS | $\hat{\pi}(x)$ CUTOFF | .0336 | .0388 | .044 | .0495 | .0555 |
| $Y_i = 1$ DEP LOSS | OBSERVED | 224 | 254 | 304 | 278 | 324 |
| | EXPECTED | 264.6 | 308.7 | 353.1 | 399.1 | 448.1 |
| | TEST STAT | 6.43 | 10 | 7.12 | 38.6 | 36.2 |
| $Y_i = 0$ ACCESSION | OBSERVED | 8319 | 8289 | 8239 | 8265 | 8219 |
| | EXPECTED | 8278 | 8234 | 8190 | 8144 | 8095 |
| | TEST STAT | .2 | .38 | .31 | 1.9 | 2.0 |

| | | LEVELS OF RISK | | | | |
|---|---|---|---|---|---|---|
| | | 17 | 17 | 16 | 14 | 10 |
| | # OF CASES | 8543 | 8543 | 8543 | 8543 | 8543 |
| ACCESSION STATUS | $\hat{\pi}(x)$ CUTOFF | .062 | .0696 | .0787 | .0903 | .1067 |
| $Y_i = 1$ DEP LOSS | OBSERVED | 340 | 402 | 470 | 562 | 657 |
| | EXPECTED | 500.8 | 560.9 | 632.1 | 720.2 | 838.5 |
| | TEST STAT | 54.9 | 48.2 | 44.9 | 37.9 | 43.5 |
| $Y_i = 0$ ACCESSION | OBSERVED | 8203 | 8141 | 8073 | 7981 | 7886 |
| | EXPECTED | 8042 | 7982 | 7911 | 7823 | 7704 |
| | TEST STAT | 3.6 | 3.4 | 3.6 | 3.5 | 4.7 |

| | | LEVELS OF RISK | | | | |
|---|---|---|---|---|---|---|
| | | 16 | 17 | 16 | 16 | 20 |
| | # OF CASES | 8543 | 8543 | 8543 | 8543 | 8545 |
| ACCESSION STATUS | $\hat{\pi}(x)$ CUTOFF | .1304 | .1707 | .2524 | .6144 | 1.0 |
| $Y_i = 1$ DEP LOSS | OBSERVED | 869 | 1125 | 1626 | 3097 | 7537 |
| | EXPECTED | 1007 | 1269 | 1759 | 3230 | 6588 |
| | TEST STAT | 21.4 | 19.1 | 12.6 | 8.86 | 597 |
| $Y_i = 0$ ACCESSION | OBSERVED | 7674 | 7418 | 6917 | 5446 | 1008 |
| | EXPECTED | 7536 | 7274 | 6784 | 5313 | 1957 |
| | TEST STAT | 2.8 | 3.3 | 3.3 | 5.3 | 2010 |

The Hosmer-Lemeshow goodness of fit statistic $\hat{C}$ is defined as follows:

$$\hat{C} = \sum_{i=1}^{20} \frac{(OBSERVED_i - EXPECTED_i)^2}{EXPECTED_i \, (1 - \bar{\pi}_i)}$$

$$WHERE, \quad \bar{\pi}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \hat{\pi}_j \quad ; \; i=1,\ldots,20 \quad ; \; j=1,\ldots,N_i$$

$$WITH \; N_i = NUMBER \; OF \; CASES \; IN \; GROUP \; i$$

Hosmer and Lemeshow demonstrated that if the fitted logistic regression model is the correct model then $\hat{C}$ has an approximate $\chi^2$ distribution with 20 - 2 = 18 degrees of freedom. The critical value, $\chi^2_{(df=18)}(\alpha = .05)$ is 28.87. The group's contributions to the test statistic $\hat{C}$ are displayed in Table X. These sum to a number much greater than 28.87. This indicates our model has significant lack of fit. An advantage of a summary test statistic like $\hat{C}$ is that it provides insight into the models fit over the 20 levels of DEP loss risk [Ref. 8:p. 144]. This model appears to fit reasonably well for those individuals that access ($y_i$ = 0) in all groups except the bottom 10% ( first two groups) and the top 5% constituting the twentieth group. Though the model in its entirety does not fit well as measured by $\hat{C}$, there appears to be potential for using its relatively good fit in all of the groups, except for these extreme groups, for predictive purposes.

Figure 6 illustrates how this misfit in the first two, and the last group impacted the value of $\hat{C}$ leading to rejection of

the hypothesis of model fit. With a perfect model, the 20 group means of the estimated probabilities of DEP loss, $\hat{\pi}$ would equal the corresponding relative frequencies of the numbers of observed values of DEP loss ($y_i$ = 1), to within random error. This would be represented by the line y = x. The curve corresponding to the fitted model appears to differ from the line y=x only for the extreme groups.

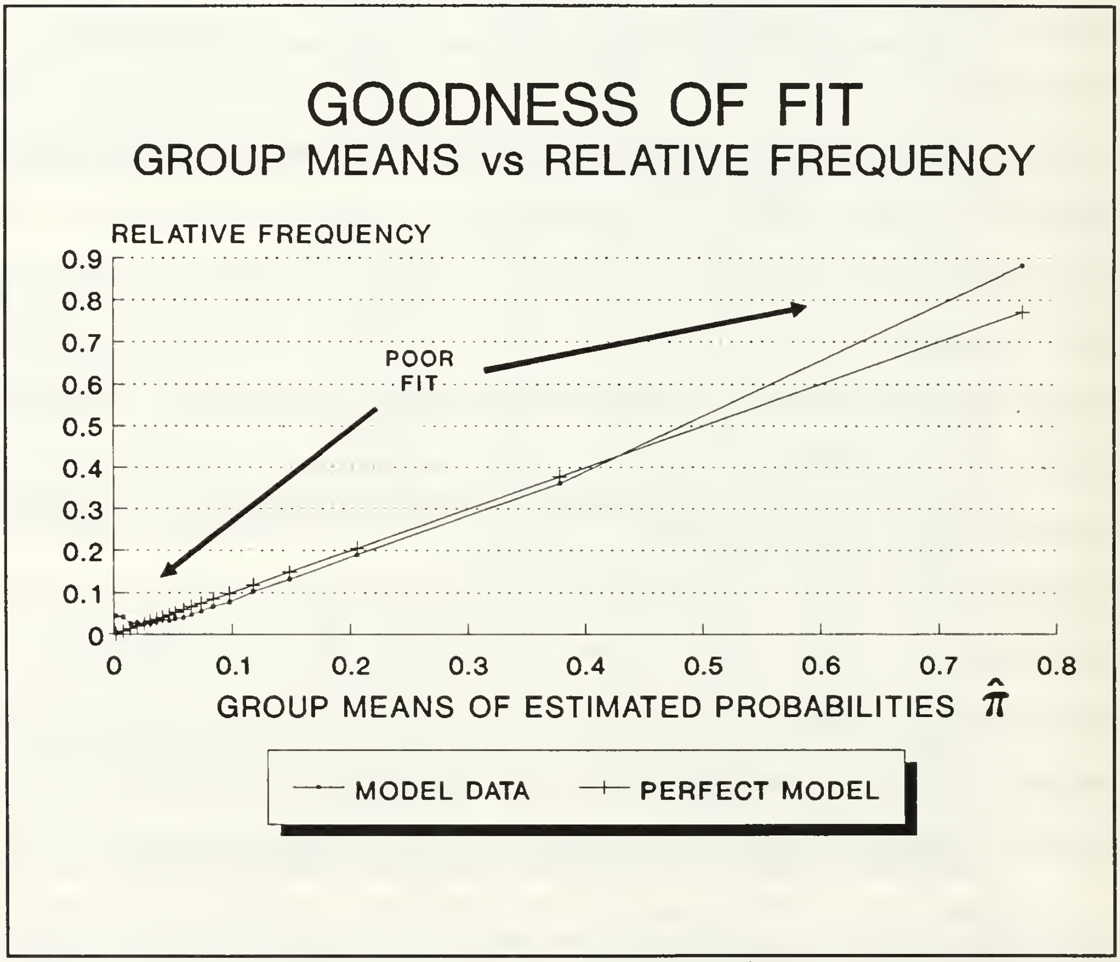


Figure 6 Hosmer-Lemeshow Goodness of Fit Plot

## C. PREDICTION PLOT

An intuitive, alternative method for assessing the fit of the developed model is the prediction plot. Figure 7 shows smoothed histograms of the estimated probability of DEP loss, $\hat{\pi}$, for both the accession and DEP loss groups. The curve for the accession group is the plot of residuals; that for the DEP loss group is a plot of one minus the residuals. Relative frequencies were plotted due to the large quantity of accession cases in comparison to the number of DEP losses.

The developed model's lack of fit is evident in the rise of the DEP loss curve to the left of $\hat{\pi} = .4$ and the low values of the same curve on the extreme right. The large area under the DEP loss curve in the region of $.6 \leq \hat{\pi} \leq .9$ appeared to indicate that the model fit well for conditions giving estimated DEP loss probabilities in this region. However, the curve for accessions indicates the model accurately classified those that accessed. As desired, the majority of those that accessed were assigned a probability of DEP loss, $\hat{\pi}$, near zero.

Though two different statistical tests indicate that the entire model was significantly different from a perfect model, closer examination reveals that the model we developed appears to perform satisfactorily for the accession and DEP loss cases in most conditions. In the next chapter, we examine the model's effectiveness in a context of its intended use for DEP loss prediction.

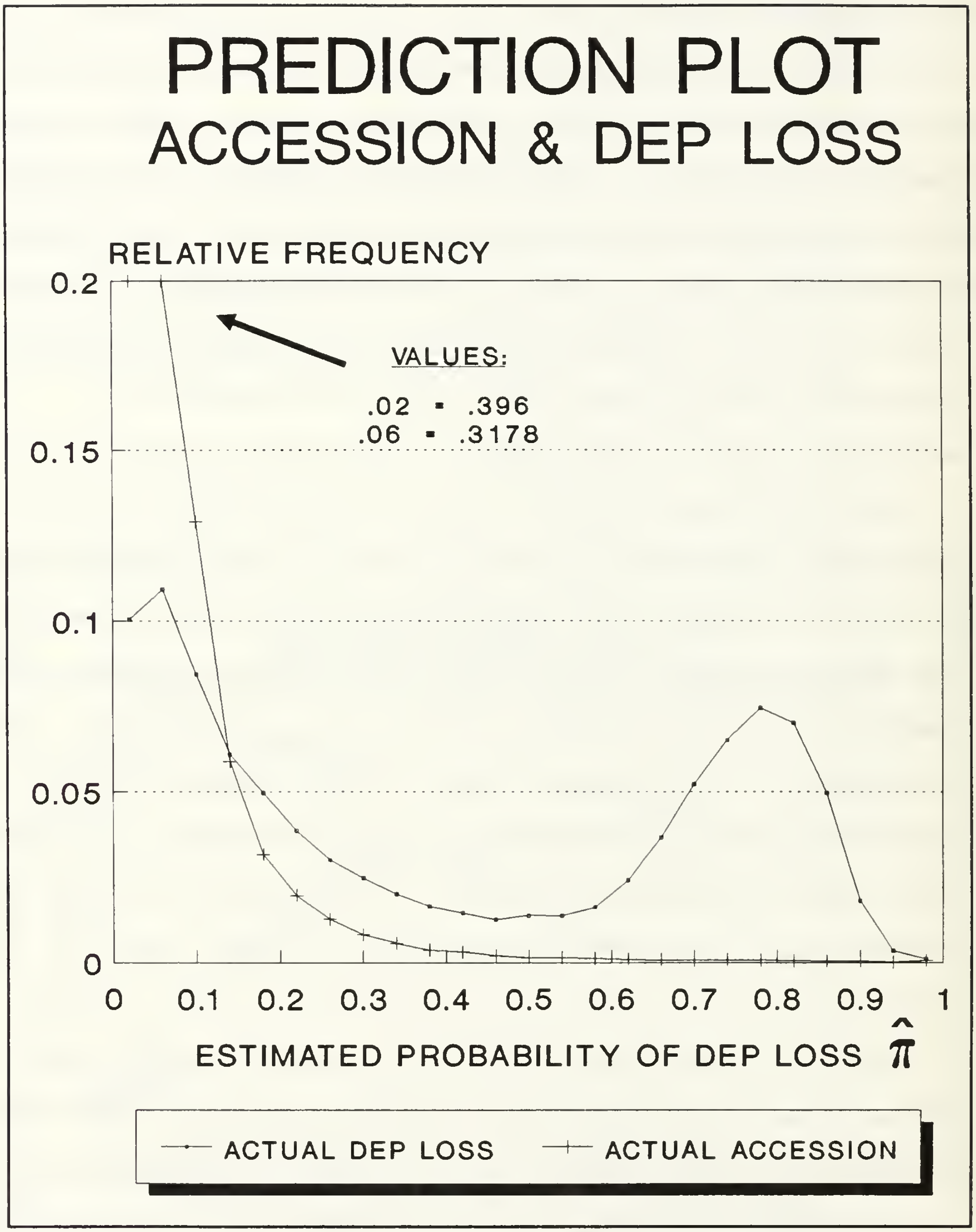

Figure 7 Prediction Plot for Accession and DEP Loss

# VII. MODEL USAGE

## A. RED, AMBER, GREEN

### 1. Classification Criteria

As mentioned in Chapter I, USAREC currently uses a red, amber, green coding scheme for recruiters to classify their DEP members according to perceived DEP loss risk. This model could provide a similar classification, augmenting the recruiters first hand knowledge of DEP members. This could prove especially helpful in classifying newly contracted DEP members, before the recruiter develops a relationship with the DEP member.

By computing and adjusting two threshold values of $\hat{\pi}$, we can control which of these three groups a DEP member is assigned. In determining these threshold values of $\hat{\pi}$, we used the following criteria. No more than one half of the DEP members would be placed in the amber group. This group is made up of the DEP members that the threshold rule will not classify as a predicted DEP loss or accession. The utility of the rule would be in question if it placed an unusually large number of DEP members in this group. USAREC could easily change this restriction on the proportion classified amber by adjusting the threshold values. The second criterion was to

maximize the model's accuracy in the classification of DEP members into the red and green categories.

### 2. GREEN Classification

The classification of a DEP member as green by the threshold value would alert the recruiter that this individual is not predicted to be a DEP loss. Figure 8 illustrates the power of the model with respect to the green category. We determined the predictive power of the fitted model is best represented by its accuracy of prediction. The predictive power for the green category increased significantly as the percentage of the total population classified green declined. Since approximately 88% of the model population accessed, an accuracy of 88% would have been achieved if all DEP members were classified as green. The power curve begins to flatten out as it approaches 50% classification green and rises no higher than 96.8% accurate at about 45% classification green. We decided to use the slightly smaller accuracy of 96.7% due to the significantly larger classification rate of 53.6% green.

As indicated in Figure 8, the cutoff threshold to maximize green classification accuracy was determined to be $\hat{\pi}(x) \leq .06$. A high accuracy is desired in the green classification because a misclassification might result in a DEP member not receiving needed extra recruiter attention.

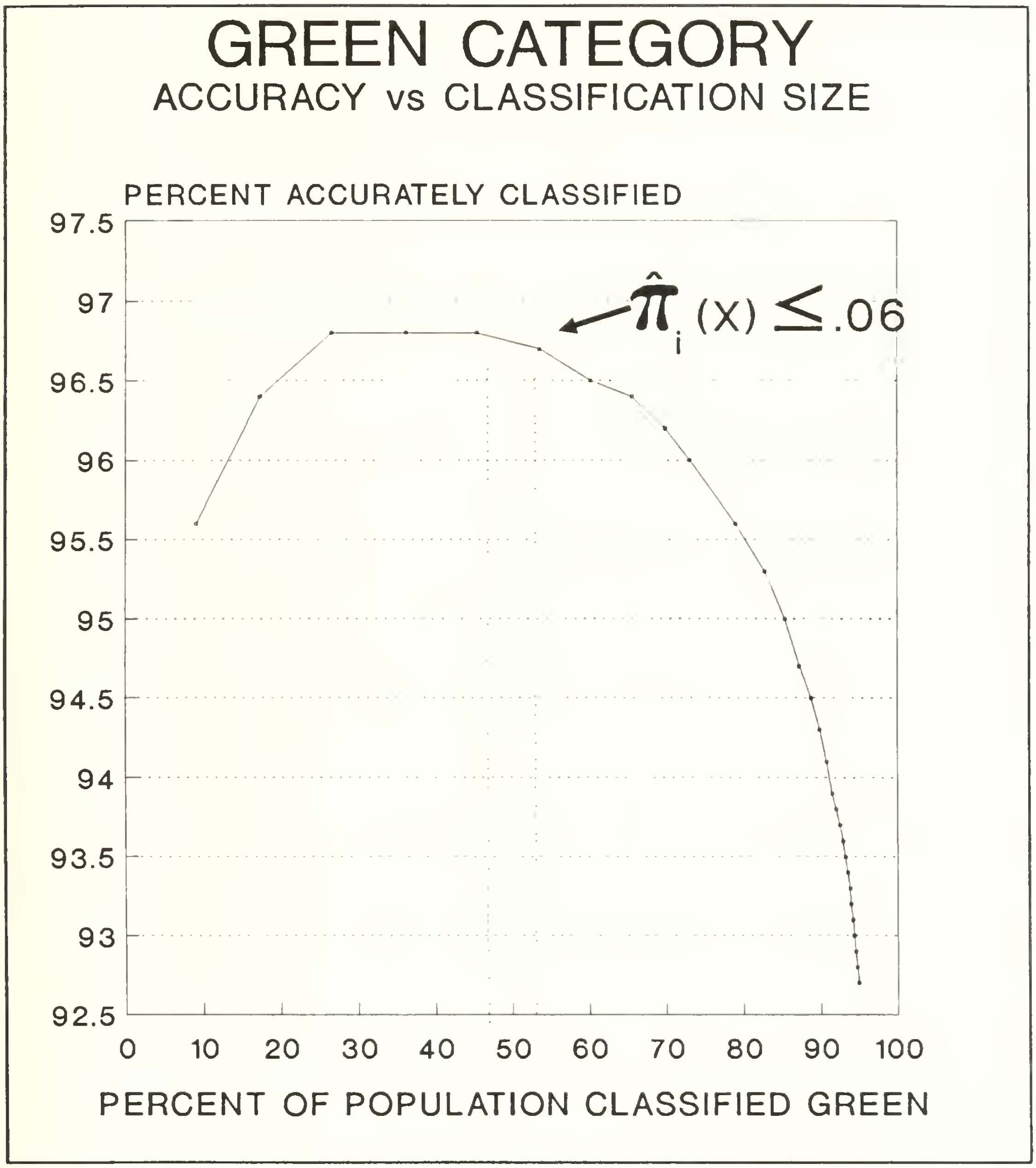


**Figure 8** Model Power Green Classification

### 3. RED Classification

The classification of a DEP member as red by the threshold value would alert the recruiter that this individual

is a high DEP loss risk and predicted to be a DEP loss. Similar to the green classification's plotted power, Figure 9 illustrates the predictive power of the fitted model with respect to classification into the red category. As in the power of the green classification, the accuracy significantly improved as the percent of the population classified red decreased. The accuracy peaked at 89.6% with a classification of about 4% of the population as red.

Though this accuracy is not as high as that of the green classification, this still appears to be a strong prediction accuracy due to the small percentage (12%) of the population that eventually became a DEP loss. For comparative purposes, the accuracy would have been only about 12% if 100% of the population was classified red. Additionally, an error in this prediction may only result in a recruiter paying closer attention to a DEP member who may have accessed without the attention. As indicated in Figure 9, the cutoff threshold used to maximize the accuracy of those classified red was $\hat{\pi}(x) \geq .70$.

### 4. Final Results

As a result of the selection of these thresholds, the final model classified the data used to fit the model as depicted in Table XI. This table indicates that less than 50% (42.45%) of the population was classified as amber. As previously mentioned, the classification accuracy was strong

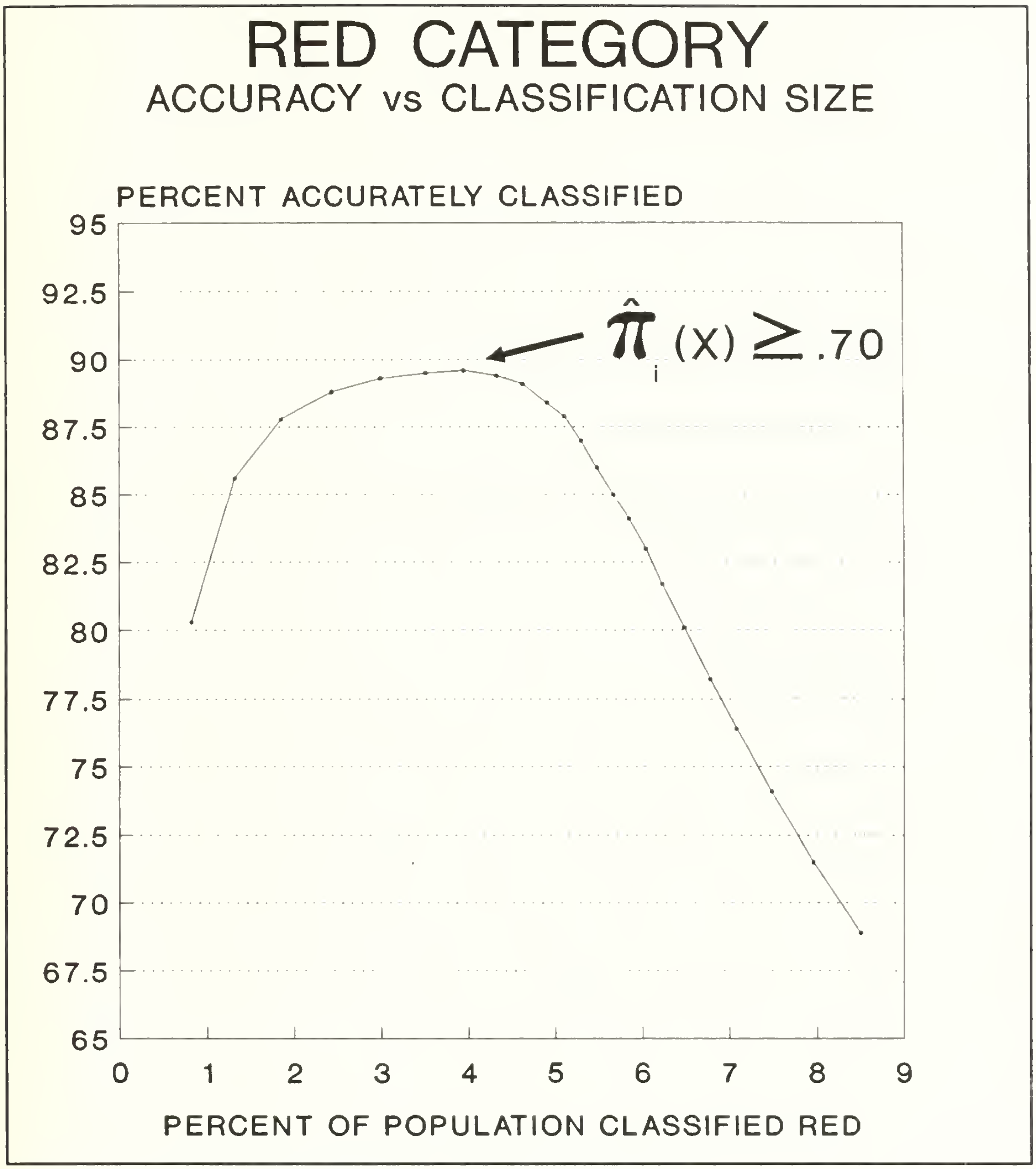

**Figure 9** Model Power Red Classification

even when constrained by no more than 50% being classified as amber. The over-all classification accuracy of the threshold rule for those DEP members that eventually did access was 99.2%; it was 66.6% for those that were DEP losses.

Table XI MODEL DATA CLASSIFICATION TABLE

| | GROUP / CRITERIA / (PERCENT OF POPULATION) | | | |
|---|---|---|---|---|
| OBSERVED | GREEN $\pi_i(x) \le .06$ ( 53.6 % ) | AMBER $.06 < \pi_i(x) < .7$ ( 42.45 % ) | RED $.7 \le \pi_i(x)$ ( 3.95 % ) | PERCENT[1] CORRECT BY $Y_i$ |
| $Y_i = 0$ ACCESSION | 88,544 | 62,141 | 704 | 99.2 % |
| $Y_i = 1$ DEP LOSS | 3,033 | 10,404 | 6,039 | 66.6 % |
| PERCENT CORRECT BY GROUP | 96.7 % | | 89.6 % | |

**NOTE:** 1. The calculation for correct by $Y_i$ does not include those classified as amber.

## B. VALIDATION

The final test conducted was the validation of the fitted model on a new data set. The method of maximum likelihood ensured that the coefficients in $\hat{B}$ were estimated so as to make the observed cases in the model data set as likely as possible. Hence, it was expected that the fitted model would perform in an optimistic manner on the model data set. Regression models with many explanatory variables at times become overly reliant on the data used to fit the model by selecting as significant, covariate patterns unique to the model data set. [Ref. 8:p. 171]

The original data set that was used to fit the model was a random 25% sample (170,685 cases) from the database of all enlistment contracts signed in FY 86 through FY 90. The new data set used for validation of the fitted model was a new

random 25% sample (171,809 cases) from the same database. Validation was conducted by calculating the logit, g(X) using the estimated coefficients from the fitted model, $\hat{B}$ in a linear combination with the covariates from the new 25% sample in equation (3). These values were then substituted into the logit transformation, equation (2), resulting in corresponding estimated probabilities, $\hat{\pi}$.

Figures 10 and 11 illustrate the predictive power of the model on a new data set as compared to the model data set. The power of the green classification on the validation data set was almost as strong as for the model data set. The maximum accuracy is obtained at the same $\hat{\pi}$ threshold with less than a .1% decrease in accuracy.

Likewise, the model performed well with the validation data set in red classification. As Figure 11 illustrates, the predictive power of the model on the validation data set was almost identical to that for the model data set. The validation data set resulted in higher prediction accuracies than the model data set when lower percentages of the validation data set were classified red.

The results of the validation effort indicate that the model is not overly reliant on the model data in either green or red classifications. Table XII summarizes the final classification results for the validation data set. Only a slightly larger percentage of individuals were classified as amber using the validation data set, still less than the

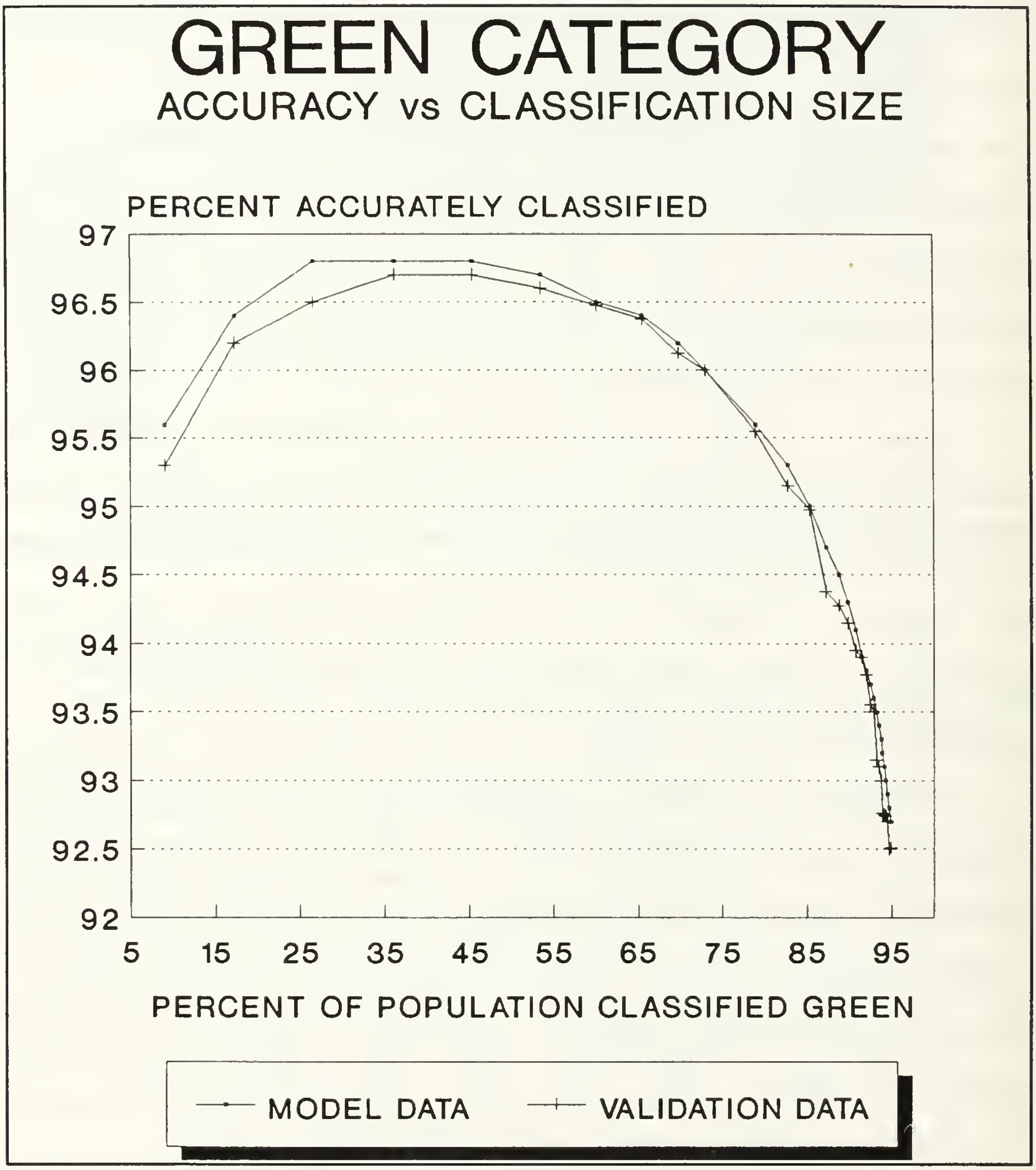


Figure 10 Green Validation Power

criterion of 50%. The red and green classification accuracies for the validation data set are only marginally smaller than the model data set. These results indicate that our model has excellent potential for predicting DEP loss outcomes for future DEP members.

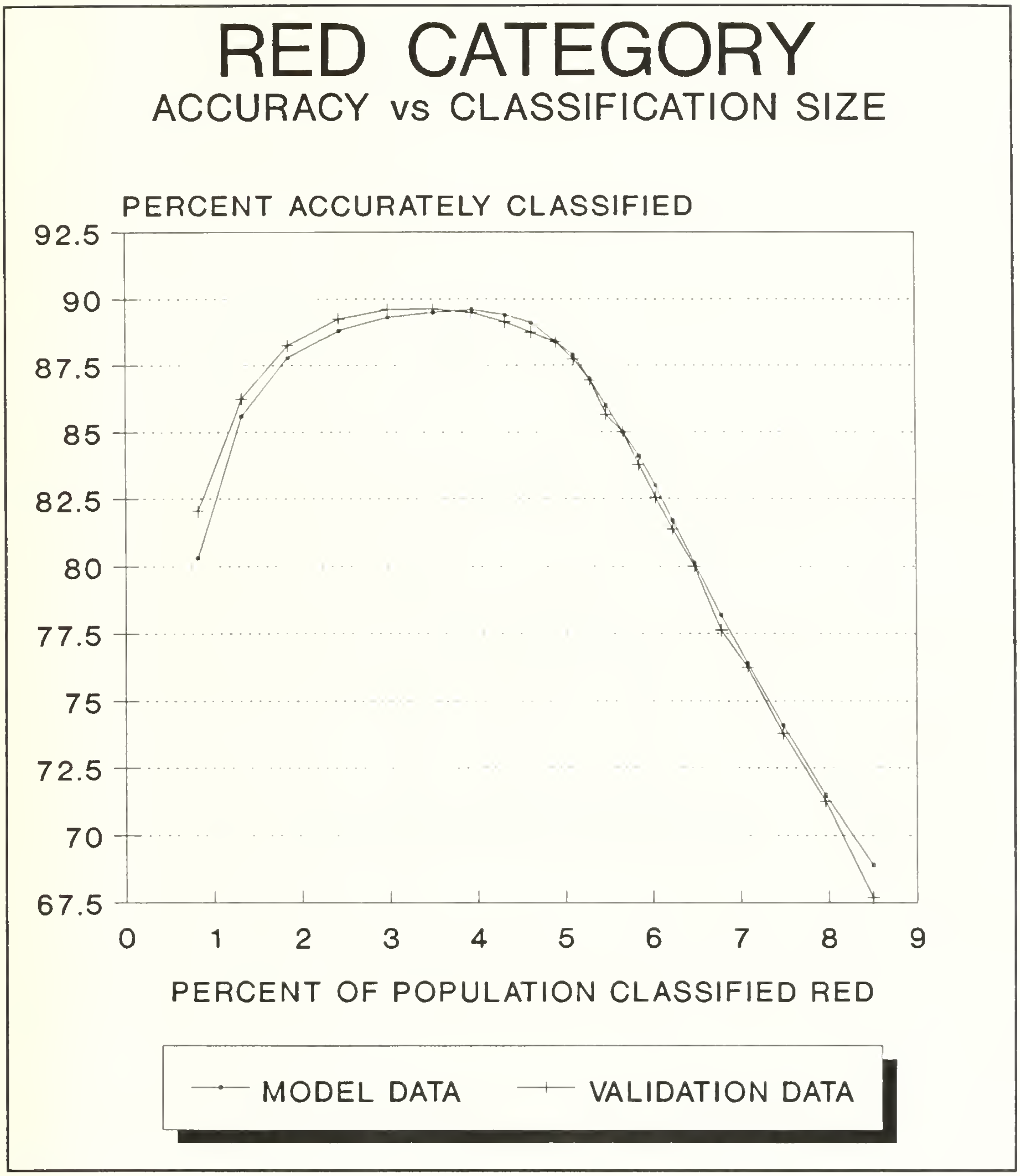

Figure 11 Red Validation Power

Table XII VALIDATION DATA CLASSIFICATION TABLE

| OBSERVED | GROUP / CRITERIA / (PERCENT OF POPULATION) GREEN $\pi_i(x) \le .06$ ( 52.9 % ) | AMBER $.06 < \pi_i(x) < .7$ ( 43.22 % ) | RED $.7 \le \pi_i(x)$ ( 3.89 % ) | PERCENT[1] CORRECT BY $Y_i$ |
|---|---|---|---|---|
| $Y_i = 0$ ACCESSION | 87,795 | 63,584 | 698 | 99.2 % |
| $Y_i = 1$ DEP LOSS | 3,070 | 10,667 | 5,995 | 66.13 % |
| PERCENT CORRECT BY GROUP | 96.62 % | | 89.57 % | |

NOTE: 1. The calculation for correct by $Y_i$ does not include those classified as amber.

# VIII. RECOMMENDATION AND CONCLUSIONS

## A. RECOMMENDATIONS

Modeling human behavior is a difficult process because there are so many unknown and unmeasurable factors which ultimately affect the dependent variable being modeled. Modeling of the DEP loss process is no exception. Therefore, recommendations that follow focus on obtaining data that could possibly act as significant explanatory variables in a refined DEP loss model.

The RENO variable used in this study indicated whether the enlistment contract had been renegotiated while the recruit was in the DEP. Though obtainable through indirect means, the USAREC Minimaster database does not describe the renegotiation process beyond a binary (yes,no) variable. Whether the renegotiation was a date change, training change, or job change might be significant information.

National and local advertising have long been considered key recruiting tools by USAREC. Analysts at USAREC have been asked in the past to quantitatively demonstrate the relationships between advertising expenditures and successful recruiting operations. The NATADV and BDEADV variables used in this fitted model were aggregated to the fiscal year. These advertising variables were not for a specific media type such

as television, radio, or newspaper. More detailed, historical advertising information down to the Recruiting Battalion level by time and media type could be valuable in developing a refined DEP loss model.

USAREC uses promotion incentives such as the E-2 referral program. DEP members who refer candidates which later sign a contract are rewarded with an advanced promotion to E-2 upon entering active duty. This has proven to be a valuable recruiting tool with respect to generating contract leads. The effect that this program may have on the DEP loss process was not modeled here due to inaccessibility of the data. Inclusion of this information in the USAREC Minimaster database could significantly assist in development of an improved DEP loss model.

## B. CONCLUSIONS

This modeling effort has attempted to quantify the complex DEP loss process involving many known explanatory variables. Though the model in its entirety did not fit well as measured by two statistical tests, for certain levels of estimated probability of DEP loss, $\hat{\pi}$, the model appeared to fit well. An important test of any model that might be used for predictive purposes is its validation. We demonstrated that our model performed satisfactorily on a validation data set obtained by taking a new 25% random sample from the database. With as an important of a resource management tool as the DEP,

a modeling effort that displays some success in predicting DEP loss should be pursued. We conclude that this model could prove useful in assisting recruiters in assessing DEP loss risks of individual recruits.

APPENDIX. A

Table XIII CAREER MANAGEMENT FIELD DEP LOSS ANALYSIS

| CLASS VARIABLE | VARIABLE DESCRIPTION | PERCENT [1] ACCESSION | PERCENT [1] DEP LOSS |
|---|---|---|---|
| CMF | CAREER MANAGEMENT FIELD | | |
| ...00 [2] | .5% CMF 00 | .6 | 99.4 |
| ...09 | .5% CMF 09 | 89.2 | 4.8 |
| ...11 | 13.8% CMF 11 | 88.4 | 10.6 |
| ...12 | 3.0% CMF 12 | 88.4 | 11.6 |
| ...13 | 7.5% CMF 13 | 80.9 | 9.9 |
| ...14 | 2.6% CMF 14 | 88.9 | 11.2 |
| ...19 | 4.3% CMF 19 | 89.7 | 10.3 |
| ...23 | .6% CMF 23 | 89.8 | 10.2 |
| ...25 | .7% CMF 25 | 88.9 | 11.3 |
| ...27 | .7% CMF 27 | 88.9 | 11.6 |
| ...29 | 1.5% CMF 29 | 88.9 | 11.1 |
| ...31 | 8.7% CMF 31 | 88.9 | 11.1 |
| ...33 | .5% CMF 33 | 89.4 | 10.6 |
| ...35 | .6% CMF 35 | 88.9 | 11.3 |
| ...46 | .1% CMF 46 | 86.0 | 14.0 |
| ...51 | 2.1% CMF 51 | 89.2 | 10.2 |
| ...54 | .9% CMF 54 | 89.2 | 9.9 |
| ...55 | .9% CMF 55 | 89.2 | 10.8 |
| ...63 | 10.4% CMF 63 | 89.1 | 10.9 |
| ...67 | 3.0% CMF 67 | 88.4 | 11.6 |
| ...71 | 5.7% CMF 71 | 86.4 | 13.6 |
| ...74 | .4% CMF 74 | 84.0 | 16.0 |
| ...76 | 7.4% CMF 76 | 88.5 | 11.5 |
| ...77 | 1.6% CMF 77 | 90.0 | 10.0 |

CONTINUED NEXT PAGE

Table XIV CAREER MANAGEMENT FIELD DEP LOSS (CONTINUED)

| CLASS VARIABLE | VARIABLE DESCRIPTION | PERCENT ACCESSION [1] | PERCENT DEP LOSS [1] |
|---|---|---|---|
| CMF | CAREER MANAGEMENT FIELD | | |
| ...81 | .8% CMF 93 | 86.7 | 13.1 |
| ...88 | 3.6% CMF 88 | 88.3 | 11.8 |
| ...91 | 7.3% CMF 91 | 86.9 | 13.1 |
| ...98 | .8% CMF 93 | 85.9 | 14.1 |
| ...94 | 3.1% CMF 94 | 88.2 | 11.8 |
| ...96 | 4.9% CMF 96 | 87.3 | 12.7 |
| ...97 | .3% CMF 97 | 94.6 | 5.4 |
| ...98 | 1.9% CMF 98 | 89.2 | 10.8 |
| TOTAL | TOTAL CONTRACT PERCENTAGES | 88.0 | 12.0 |

NOTE: 1. Cell difference significance less than .00005 Chi-square test 2. This is not real CMF but only a surrogate "holding" CMF for a known DEP loss who is not being carried on record as a DEP loss. Discussed in Chapter IV.

Table XV RECRUITING BATTALION DEP LOSS ANALYSIS

| CLASS VARIABLE | VARIABLE DESCRIPTION | PERCENT ACCESSION [1] | PERCENT DEP LOSS [1] |
|---|---|---|---|
| BATTALION | RECRUITING BATTALION | | |
| ...1A | 1.0% FROM ALBANY | 85.6 | 14.3 |
| ...1A | 2.6% FROM BALTIMORE | 87.9 | 12.1 |
| ...1G | 1.3% FROM BOSTON | 83.9 | 10.8 |
| ...1A | 1.0% FROM BRUNSWICK | 82.0 | 10.0 |
| ...1E | 1.9% FROM HARRISBURG | 86.1 | 13.9 |
| ...1F | .9% FROM NEW HAVEN | 85.9 | 14.1 |
| ...1G | 1.9% FROM NEW YORK CITY | 85.4 | 10.6 |
| ...1N | 1.4% FROM NEWBURGH | 82.8 | 17.2 |
| ...1K | 1.6% FROM PHILADELPHIA | 85.6 | 10.6 |
| ...1F | 2.2% FROM PITTSBURGH | 87.9 | 12.4 |
| ...1N | 2.0% FROM SYRACUSE | 86.9 | 11.1 |
| ...3A | 2.5% FROM ATLANTA | 88.9 | 11.1 |
| ...1A | 1.5% FROM BECKLEY | 88.4 | 11.6 |
| ...3C | 1.5% FROM CHARLOTTE | 88.9 | 11.4 |
| ...1N | 1.9% FROM COLUMBIA | 92.2 | 8.6 |
| ...3E | 2.8% FROM JACKSONVILLE | 88.9 | 11.1 |
| ...3F | 1.7% FROM LOUISVILLE | 82.8 | 11.4 |
| ...3G | 2.5% FROM MIAMI | 87.5 | 12.5 |
| ...3H | 2.3% FROM MONTGOMERY | 93.8 | 8.1 |
| ...3I | 1.7% FROM NASHVILLE | 87.9 | 12.9 |
| ...3J | 1.7% FROM RALEIGH | 91.6 | 8.6 |
| ...3K | 1.9% FROM RICHMOND | 93.8 | 8.1 |
| ...4A | 1.5% FROM ALBUQUERQUE | 83.9 | 10.3 |
| ...4E | 2.5% FROM DALLAS | 93.0 | 11.0 |
| ...4C | 1.8% FROM DENVER | 89.2 | 10.8 |
| ...4E | 2.4% FROM HOUSTON | 91.1 | 8.6 |
| ...4F | 2.1% FROM JACKSON | 85.4 | 10.6 |
| ...4G | 2.1% FROM KANSAS CITY | 88.9 | 11.1 |
| ...4I | 2.1% FROM LITTLE ROCK | 93.8 | 8.6 |
| ...4I | 1.8% FROM NEW ORLEANS | 93.8 | 6.2 |
| ...4J | 1.6% FROM OKLAHOMA CITY | 91.4 | 8.6 |
| ...4A | 2.0% FROM SAN ANTONIO | 91.9 | 8.1 |
| ...4N | 1.9% FROM ST LOUIS | 87.1 | 12.9 |

CONTINUED NEXT PAGE

**Table XVI RECRUITING BATTALION DEP LOSS (CONTINUED)**

| CLASS VARIABLE | VARIABLE DESCRIPTION | PERCENT ACCESSION [1] | PERCENT DEP LOSS [1] |
|---|---|---|---|
| BATTALION | RECRUITING BATTALION | | |
| ...5A | 2.0% FROM CHICAGO | 88.9 | 11.9 |
| ...5B | 1.4% FROM CINCINNATI | 86.3 | 13.7 |
| ...5C | 2.5% FROM CLEVELAND | 86.1 | 14.6 |
| ...5D | 1.5% FROM COLUMBUS | 88.9 | 12.0 |
| ...5F | 1.2% FROM DES MOINES | 90.6 | 14.6 |
| ...5F | 2.2% FROM DETROIT | 87.4 | 12.6 |
| ...5B | 1.8% FROM INDIANAPOLIS | 88.9 | 11.1 |
| ...5I | 2.4% FROM LANSING | 88.9 | 13.7 |
| ...5L | 1.9% FROM MILWAUKEE | 85.5 | 14.5 |
| ...5K | 1.8% FROM MINNEAPOLIS | 86.3 | 13.7 |
| ...5L | 1.7% FROM OMAHA | 90.6 | 10.4 |
| ...5A | 1.8% FROM PEORIA | 87.4 | 13.0 |
| ...6A | 1.6% FROM SAN FRANCISCO | 86.3 | 15.4 |
| ...6E | .8% FROM HONOLULU | 89.3 | 13.7 |
| ...6F | 2.9% FROM LOS ANGELES | 86.1 | 13.0 |
| ...6G | 1.8% FROM PHOENIX | 88.9 | 11.9 |
| ...6H | 1.5% FROM PORTLAND | 87.7 | 12.3 |
| ...5I | 2.0% FROM SACRAMENTO | 88.9 | 11.1 |
| ...6J | 1.4% FROM SALT LAKE CITY | 90.6 | 9.4 |
| ...6A | 2.1% FROM SANTA ANA | 87.4 | 12.6 |
| ...6L | 2.1% FROM SEATTLE | 88.5 | 11.5 |
| TOTAL | ALL CONTRACTS | 88.0 | 12.0 |

NOTE: 1. Cell difference significance less than .00005 Chi-square test

Table XVII ESTIMATED COEFFICIENTS EDUC, RACE, PADDMO

| VARIABLE | ESTIMATED COEFFICIENT $\hat{B}_i$ | SIGNIFICANCE LEVEL |
|---|---|---|
| EDUC EDUCATION STATUS AT CONTRACT | | |
| ...SENIOR | .6642 | .0004 |
| ...NONGRAD | -1.4175 | .0000 |
| ...DIPGRAD | 1.8728 | .0000 |
| ...MONGRAD | -1.1195 | .0000 |
| RACE | | |
| ...WHITE | .2521 | .0008 |
| ...OTHER | .2275 | .0774 |
| ...BLACK | .0147 | .8569 |
| ...ASIAN | -.3602 | .0000 |
| ...HISPAN | -.1341 | .0000 |
| PADDMO PROJECTED ACCESSION MONTH | | |
| .....JAN | .0954 | .0000 |
| .....FEB | -.1290 | .0004 |
| .....MAY | .0446 | .2407 |
| .....FEB | .0012 | .9777 |
| .....MAY | .1047 | .0037 |
| .....JUN | .1116 | .0000 |
| .....JUL | .0012 | .7682 |
| .....AUG | -.1290 | .0000 |
| .....OCT | -.1781 | .0000 |
| .....OCT | .0949 | .0008 |
| .....NOV | -.0124 | .6635 |
| .....DEC | -.0532 | .0000 |

Table XVIII ESTIMATED COEFFICIENTS FOR CMF

| VARIABLE | ESTIMATED COEFFICIENT $\hat{B}_i$ | SIGNIFICANCE LEVEL |
|---|---|---|
| CMF CAREER MANAGEMENT FIELD | | |
| 91 | .1052 | .012 |
| 31 | .1243 | .0000 |
| 91 | .1478 | .0000 |
| 91 | .1130 | .0008 |
| 71 | .242 | .0000 |
| 27 | .1245 | .2358 |
| 19 | .0465 | .3559 |
| 12 | .1289 | .023 |
| 76 | .1324 | .0025 |
| 91 | .3131 | .0069 |
| 94 | .383 | .0000 |
| 13 | .1167 | .0047 |
| 88 | .1186 | .0000 |
| 76 | .1289 | .0069 |
| 98 | -.1797 | .012 |
| 14 | .0683 | .2614 |
| 33 | .1186 | .0000 |
| 09 | -2.2913 | .0069 |
| 67 | .0392 | .503 |
| 09 | -.1615 | .1781 |
| 74 | .3141 | .0000 |
| 23 | .0174 | .8857 |
| 94 | .0874 | .012 |
| 29 | .1830 | .0132 |
| 98 | -1.4193 | .0000 |
| 13 | .2513 | .0069 |
| 55 | .2613 | .0063 |
| 25 | -.002 | .9827 |
| 81 | -.0351 | .8366 |
| 77 | .2045 | .0049 |
| 46 | 0.31 | .0000 |

Table XIX ESTIMATED COEFFICIENTS FOR BN

| VARIABLE | ESTIMATED COEFFICIENT $\hat{B}_i$ | SIGNIFICANCE LEVEL |
|---|---|---|
| BN RECRUITING BATTALION | | |
| 1A | -.218 | .0192 |
| 1G | 1.2767 | .0000 |
| 1G | .9065 | .0000 |
| 1G | -.1931 | .052 |
| 3F | -.3692 | .0000 |
| 1L | 1.1548 | .0000 |
| 1G | .7244 | .0000 |
| 1H | 2.2415 | .0000 |
| 1K | .7448 | .0000 |
| 1L | -.218 | .0137 |
| 1N | .2085 | .0000 |
| 3B | -.0776 | .2375 |
| 3B | -.606 | .0000 |
| 3C | -.1493 | .0672 |
| 3D | -.7523 | .0000 |
| 3F | -.4442 | .0000 |
| 3F | -.7908 | .0045 |
| 3G | .0094 | .8008 |
| 3B | -.6782 | .0000 |
| 3F | -.1921 | .0672 |
| 3J | -.1086 | .0000 |
| 3B | -.1086 | .0000 |
| 4D | -.7307 | .0000 |
| 4D | .2034 | .0034 |
| 4D | .9065 | .3445 |
| 4E | .3142 | .0000 |
| 4D | -.7908 | .0000 |
| 4G | -.1932 | .0069 |
| 4D | -1.2489 | .0000 |
| 4I | -1.2526 | .0000 |
| 4D | -.5316 | .0000 |
| 4K | -.8077 | .0000 |
| 4N | .0053 | .9408 |

Table XX ESTIMATED COEFFICIENTS FOR BN (CONTINUED)

| VARIABLE | ESTIMATED COEFFICIENT $\hat{B}_i$ | SIGNIFICANCE LEVEL |
|---|---|---|
| BN RECRUITING BATTALION | | |
| 6F | .9058 | .0000 |
| 5B | .2358 | .0045 |
| 5C | .3996 | .0000 |
| 5D | -.1517 | .0000 |
| 5C | -.0939 | .2550 |
| 5F | .8422 | .0000 |
| 5H | .0722 | .3109 |
| 5F | .0364 | .5956 |
| 5J | .1939 | .0074 |
| 5D | .3535 | .0000 |
| 5L | -.9025 | .0000 |
| 5M | .5068 | .0000 |
| 6F | 1.0243 | .0000 |
| 6E | -.5372 | .0000 |
| 6F | .2149 | .0026 |
| 6G | -.0333 | .0000 |
| 5H | -.2194 | .0000 |
| 6I | -.4852 | .0000 |
| 6J | -.5277 | .0000 |
| 6K | .3903 | .0000 |
| 6L | .080 | .0000 |

## LIST OF REFERENCES


1. US Army Research Institute for the Behavioral and Social Sciences Report 86-2, Research Overview of the US Army's Delayed Entry Program, by Jeanna F. Celeste, p. 2, January 1986.

2. United States Army Recruiting Command, Regulation 601-95, Chapter 2, 27 March 1986.

3. US Army Research Institute for the Behavioral and Social Sciences Report 679, The Delayed Entry Program: A Policy Analysis, by Alex G. Manganaris and Chester E. Phillips, p. 14, May 1985.

4. Rick Maze, "Gulf war fallout puts strain on recruiters," Army Times, p. 8, 13 May 1991.

5. US Army Research Institute for the Behavioral and Social Sciences Report 823, Delayed Entry Program (DEP) Loss Behavior, p. 1-4, September 1988.

6. US Army Research Institute for the Behavioral and Social Sciences Report 832, Estimating the Army's Prime Recruiting Market, p. vii, April 1989.

7. Department of Defense Manpower Data Center Letter Ser: 85E-10 to Distribution United States Army Recruiting Command, Subject: Supersite Demographic Data Base, 15 January 1985.

8. Hosmer, David W., and Lemeshow, Stanley, Applied Linear Regression, John Wiley & Sons, Inc., 1989.

9. SPSS, Inc., Advanced Statistical Guide, Version 4.0, Procedure Logistic Regression, SPSS, Inc., 1990.

10. Fienberg, Stephen E., The Analysis of Cross-Classified Categorical Data, 2d ed., The MIT Press, 1989.

## INITIAL DISTRIBUTION LIST

1. Commander 1
   U.S. Army Training and Doctrine Command
   ATTN: ATCD
   Fort Monroe, VA 23651-5000

2. Deputy Undersecretary of the Army 1
   for Operations Research
   ATTN: Mr. Hollis
   Room 2E660, The Pentagon
   Washington, D.C. 20310-0102

3. Library, Code 52 2
   Naval Postgraduate School
   Monterey, CA 93943-5002

4. Defense Technical Information Center 2
   ATTN: DTIC, TCA
   Cameron Station
   Alexandria, VA 22304-6145

5. U.S. Army Library 1
   Army Study Documentation and Information Retrieval System
   ANRAL-RS
   ATTN: ASDIRS
   Room 1A518, The Pentagon
   Washington, D.C. 20310

6. Professor Donald Barr 1
   Department of Mathematics
   Naval Postgraduate School, Code MA / Ba
   Monterey, CA 93943-5000

7. Captain Daniel C. Buning, USA 1
   1228 Lake Williesara Cir.
   Orlando, Fl 32806

8. United States Army Recruiting Command 2
   Program Analysis and Evaluation Directorate
   Fort Sheridan, IL 60037

Thesis
B852 Buning
c.1 US Army's Delayed Entry
Program : attrition
modeling.

DUDLEY KNOX LIBRARY
3 2768 00031913 1